Spatial patterns in shallow cumulus cloud populations over a heterogeneous surface

INAUGURAL-DISSERTATION

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Thirza W. van Laar

aus

Barneveld

Die Niederlande

Köln 2019

Berichterstatter:

Prof. Dr. Roel A.J. Neggers
Prof. Dr. Clemens Simmer

Tag der letzten mündlichen Prüfung:

22 July 2019

Rows and flows of angel hair
And ice cream castles in the air
And feather canyons everywhere
I've looked at clouds that way

But now they only block the sun
They rain and snow on everyone
So many things I would have done
But clouds got in my way

I've looked at clouds from both sides now
From up and down, and still somehow
It's cloud illusions that I recall
I really don't know clouds at all

Joni Mitchell - Both Sides Now

Contents

Abstract

Shallow cumulus clouds play an important role in the earth's climate. They are effective transporters of temperature, humidity and momentum and impact the radiation budget. It is therefore important to correctly represent them in numerical weather prediction and climate models. However, because of their small scale they are not directly resolved by the large scale models. Their presence and effect therefore is approximated by parametrization schemes. These parametrization schemes use many assumptions and have their uncertainties, therefore clouds are one of the major sources of uncertainty for climate prediction.

To minimize the uncertainties associated with the different parametrization schemes, a more complete understanding of what influences cloud formation is needed. In this thesis two specific aspects of processes associated with shallow cumulus cloud populations are studied. The first one is their spatial organization. Through convective organization the mean state of the atmosphere is affected. Quantifying this behaviour could help in understanding the mechanisms behind organization. The second one is their dependence on surface conditions, more in particular the role of surface heterogeneity. Since shallow cumulus are strongly coupled to the surface, their formation and spatial distribution is influenced by heterogeneous surface conditions. In this thesis first some exploratory work is done on the description of a shallow cumulus cloud population in terms of size and spacing. The tools acquired are applied to asses the influence of a heterogeneous surface on the spatial patterns in shallow cumulus cloud populations.

For a statistically reliable assessment of cloud size distributions, Large Eddy Simulations (LES) are used. 146 simulations are performed for days that feature shallow cumulus clouds. It is found that the cloud size distribution can be described by a power law-exponential function. The largest cloud in the field is found to correlate with the total cloud cover, meaning that larger clouds contribute most to a larger total cloud cover, and not an increase in number of smaller ones.

To study cloud spacing, data is used from a large domain LES over the ocean in the subtropics. The data shows that the more numerous small clouds surround larger ones. The distances between clouds depend on the size of the cloud itself. The larger the cloud, the larger the distance to its nearest neighbor. The functional relation between cloud spacing and cloud size differs when either all clouds are taken into account, or only clouds of a similar size. For clouds of a similar size the spacing is found to increase exponentially with cloud size. To quantify the degree of organization of a complete cloud population, several parameters are evaluated. Taking into account advantages and disadvantages of all, it is concluded that I_{org} is the best one to use, mainly because it describes the tendencies seen by eye and is useful over a range of scales.

Finally the impact of surface heterogeneity on the cloud population is studied. For this, two different shallow cumulus days are simulated using a realistic set-up with cloud resolving resolutions. A sensitivity study is done, increasing and decreasing the topography, and changing the distribution of land use types. The cloud size distribution is not greatly affected by these changes in surface conditions. The slope stays the same, only the maximum

cloud size differs slightly for the simulations, with the largest clouds for the simulations with increased topography. Judging by eye, the spatial distribution of clouds differs among the simulations, but this is not reflected in I_{org}. For the simulation with a different distribution of land use types a quasi secondary circulation can form if the wind direction allows for it. Even though quantification is not straightforward, using a realistic set-up shows that surface conditions do influence the spatial patterns in shallow cumulus cloud populations.

Zusammenfassung

Flache Kumuluswolken spielen eine wichtige Rolle im Klimasystem. Sie transportieren Temperatur, Feuchte und Impuls und beeinflussen die Strahlungsbilanz. Deshalb ist es wichtig sie korrekt in numerischen Wettervorhersage- und Klimamodellen zu repräsentieren. Allerdings werden sie aufgrund ihrer kleinen Skalen nicht direkt von den großskaligen Modellen aufgelöst. Ihr Erscheinen und ihr Einfluß werden deshalb durch Parameterisierungen angenähert. Diese Parameterisierungen nutzen viele Annahmen und haben eigene Unsicherheiten, weshalb Wolken eine der Hauptquellen von Unsicherheiten in Klimavorhersagen sind.

Um die Unsicherheiten der verschiedenen Parameterisierungen zu minimieren wird ein umfassenderes Verständnis der Einflüsse auf die Entstehung von Wolken gebraucht. In dieser Arbeit werden zwei spezifische Aspekte von Prozessen untersucht, die mit flachen Kumuluswolken Populationen verbunden sind. Der erste Aspekt ist die räumliche Organization. Durch konvektive Organisation wird der mittlere Zustand der Atmosphäre beeinflusst. Dieses Verhalten zu quantifizieren könnte helfen die Mechanismen hinter der Organisation zu verstehen. Der zweite Aspekt ist die Abhängigkeit von der Oberflächenbeschaffenheit, spezifischer der Rolle der Oberflächenheterogenität. Da flache Kumuluswolken stark mit der Oberfläche gekoppelt sind, wird ihre Entstehung und räumliche Verteilung durch die Heterogenität der Oberflächenbeschaffenheit beeinflusst. In dieser Arbeit wird zunächst erforscht, wie Populationen von flachen Kumuluswolken anhand von Größe und Abstand charakterisiert werden können. Die erworbenen Werkzeuge werden dann genutzt um den Einfluss der heterogenen Oberfläche auf die räumlichen Muster der Populationen von flachen Kumuluswolken zu beurteilen.

Für eine zuverlässige statistische Bewertung von Wolkengrößenverteilungen, werden Large-Eddy-Simulationen (LES) verwendet. Insgesamt werden 146 Simulationen für Tage mit flachen Kumuluswolken durchgeführt. Es zeigt sich, dass die Wolkengrößenverteilung mit einer Potenzgesetz-exponentiellen Funktion beschrieben werden kann. Dabei korreliert die größte Wolke in dem Feld mit der totalen Wolkenbedeckung, das heißt, dass größere Wolken am meisten zu einer erhöhten totalen Wolkenbedeckung beitragen und nicht eine Erhöhung der Anzahl kleinerer Wolken.

Um den Wolkenabstand zu untersuchen, werden Daten von einer LES über einem großen Gebiet über dem subtropischen Ozean genutzt. Die Daten zeigen, dass die zahlreicheren kleinen Wolken die großen Wolken umrunden. Der Abstand zwischen Wolken hängt von der Größe der Wolken ab. Je größer die Wolke ist, desto größer ist der Abstand zum nächsten Nachbarn. Die funktionale Beziehung zwischen Wolkenabstand und Wolkengröße ändert sich, abhängig davon ob alle Wolken oder nur Wolken einer ähnlichen Größe berücksichtigt werden. Es zeigt sich, dass für Wolken einer ähnlichen Größe der Abstand exponentiell mit der Größe der Wolken ansteigt. Um den Grad der Organisation einer kompletten Wolkenpopulation zu quantifizieren werden verschiedene Parameter ausgewertet. Wenn man die Vor- und Nachteile von allen berücksichtigt, zeigt es sich, dass I_{org} am besten geeignet ist,

hauptsächlich weil es die mit bloßem Auge zu erkennenden Tendenzen beschreibt und es über einem Interval an Skalen anwendbar ist.

Zum Abschluß wird der Einfluß der Oberflächenheterogenität auf die Wolkenpopulation untersucht. Hierfür werden zwei verschiedene Tage mit flacher Kumulusbewölkung simuliert unter Anwendung eines realistischen Setups mit wolkenauflösender Auflösung. Eine Sensitivitätsstudie wird durch Verstärkung und Verringerung der Topographie und Änderung der Verteilung der Landnutzungstypen durchgeführt. Die Wolkengrößenverteilung wird durch diese Änderungen der Oberflächenbeschaffenheit nicht großartig beeinflusst. Die Neigung bleibt gleich und nur die maximale Wolkengröße unterscheidet sich leicht während der Simulationen, wobei die größte Wolke in den Simulationen mit verstärkter Topographie gefunden wird. Mit bloßem Auge betrachtet, scheint sich die räumliche Verteilung der Wolken in den Simulationen zu unterscheiden, aber dies spiegelt sich nicht in I_{org} wieder. In den Simulation mit verschiedenen Verteilungen von Landnutzungstypen kann sich eine quasi sekundäre Zirkulation ausbilden, wenn die Windrichtung dies zulässt. Obwohl eine Quantifizierung nicht einfach möglich ist, zeigt es sich durch die Nutzung eines realistischen Setups, dass die Oberflächenbeschaffenheit die räumlichen Muster von Populationen von flachen Kumuluswolken beeinflusst.

Chapter 1

Motivation

1.1 Shallow Cumulus clouds

Shallow cumulus clouds (ShCu) occur in abundance and cover large parts of the earth. They can be characterized as fluffy and patchy, with large areas of blue sky in between them. For this reason their other name is fair-weather clouds. A simple search online gives many hits where these characteristics can be seen, a few examples are shown in Figure 1.1. They form in the layer of the atmosphere closest to the earth, this layer is called the boundary layer. Other types of clouds that form in the boundary layer are stratocumulus and deep convection. ShCu distinguish themselves from stratocumulus by their patchiness (stratocumulus has an extensive deck with high cloud covers) and from deep convection by their shallowness and by the fact that they usually do not precipitate.

ShCu can be frequently observed, especially in the subtropics (Eastman et al., 2011) where due to large scale circulations they can be found year round. ShCu also occur over land in the mid-latitudes. Here they mainly form during the summer months, since for their formation sufficient energy at the surface is necessary. Independent of where they are present, ShCu influence the temperature and moisture distribution in the boundary layer and impact the radiation budget. For a better understanding of the impact they have on the atmosphere it is important to understand the processes related to their formation.

1.2 Impact of shallow cumulus on the atmosphere

In the formation process of ShCu, heat and moisture are transported from the surface to the boundary layer, thereby heating and moistening the air higher aloft. In the cloud layer they are the source of turbulence and convection. Through the mixing of air they initiate a redistribution of temperature and moisture.

Besides their influence on the temperature and humidity distribution, ShCu also influence the radiation budget. Because of their high albedo, clouds reflect solar radiation back to space. Accordingly, less radiation reaches the surface which has a cooling effect on the atmosphere. At the same time clouds emit longwave radiation back to the surface, thereby trapping the energy in the boundary layer. This has a warming effect on the atmosphere. Which of these two effects is stronger depends on the type of cloud and the specific properties of the clouds. These specific properties include height, location, size, thickness and liquid water content. Stratocumulus for example, with their extensive cloud decks, have a different effect than patchy shallow cumulus. The total effect of clouds on the radiation budget is called the cloud radiative forcing (CRF). A difference in cloud

Figure 1.1: Some examples of shallow cumulus clouds.
Sources: https://sciencestruck.com/cumulus-clouds-information
https://www.ktbs.com/news/arklatex-indepth/all-about-the-clouds
https://sciencestruck.com/cumulus-clouds-facts
http://exchange.smarttech.com/details.html?id=356a87dc-aa46-485d-8b65-5e91fe6bcf66

thickness, amount of clouds or cloud size over time can change the total CRF. A difference in CRF will in turn again impact cloud formation, and so on and so forth. This feedback loop is called the cloud feedback.

1.3 Clouds in a changing climate

In the future, our climate is predicted to change. Due to increased CO_2 levels in the atmosphere, an increase in sea surface temperature is expected. This will lead to higher temperatures in the whole atmosphere, causing higher levels of water vapour as well. This will change the radiation of the atmosphere. Global circulations will change, which will impact horizontal wind speeds and advection of temperature and moisture. All these changes will influence the formation of clouds.

In the current climate the CRF leads to a cooling of the atmosphere. As a response to increased sea surface temperatures this might change. The cooling might be weakened, or even turn into warming. The sign of the cloud feedback and the impact of clouds on climate in general therefore is a process that is important to understand. To study this, climate models are used. These are global, large scale simulations that operate on long time scales. Using 12 different climate models, the increase in temperature is coupled to different feedbacks (Dufresne and Bony 2008, Fig. 1.2). These are the water vapour feedback, surface albedo feedback, cloud feedback and Planck response. For all feedbacks except the cloud feedback, the temperature change is similar for all models. The cloud feedback however shows a large spread among the models. This shows that the cloud feedback is a major source of uncertainty in climate models (Vial2013). Determining cloud cover and the CRF is difficult, because of the small scale of the processes involved and the large scales of the models. However, small changes in the representation of clouds can lead to big effects because of the feedback loops associated with it. A correct representation of clouds in climate models is therefore required. This is the reason the World Climate Research Programme (WCRP) initiated a grand challenge on Clouds, Circulation and Climate Sensitivity.

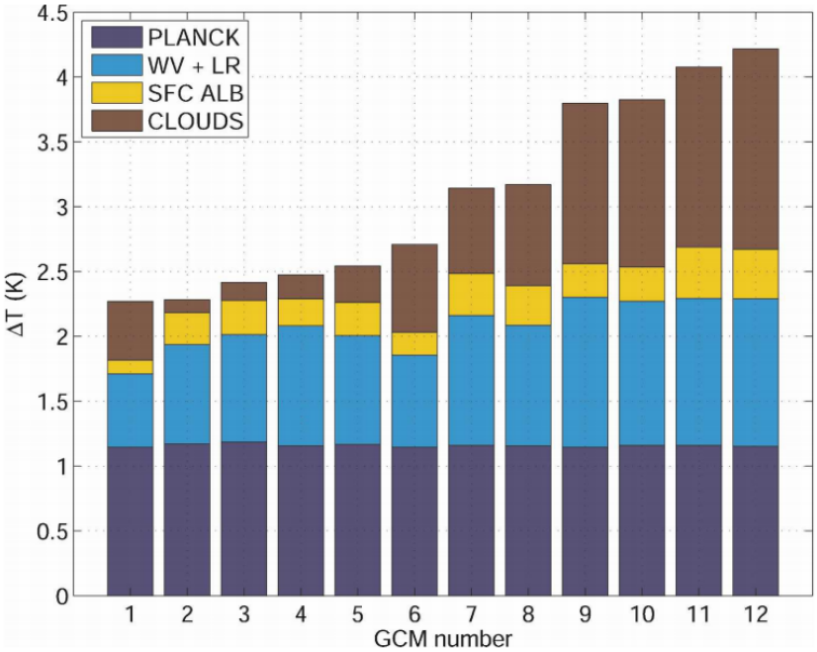

Figure 1.2: Effect of different feedbacks on temperature increase as determined with 12 different climate models. The variation between the models is largest for the cloud feedback. The water vapour feedback is referred to as WV+LR, SFC ALB indicates the surface albedo feedback (Dufresne and Bony, 2008).

1.4 Processes of importance for cloud formation

For the improvement of the representation and impact of ShCu in climate models to reduce the uncertainty related to them, the processes associated with their formation need to be better understood. To understand a process, it is helpful to first be able to describe or to quantify that what can be observed. In this thesis, the focus lies on describing and quantifying two processes that influence cloud formation. These are convective organization and the impact of surface heterogeneity.

1.4.1 Convective organization

Clouds are affected by their environment because of the mixing of air the cloud edges. It is then through the environment that they affect each other and interact. How the interaction influences the sizes of the clouds and the cloud spatial distribution is a topic of research. In idealized simulations as well as satellite observations of small to meso-scale convection, a clustering or organization of clouds is observed, although the mechanisms for this are not fully understood yet (Wing, 2019). Convective organization has been shown to influence the mean characteristics of the boundary layer (Wing and Cronin, 2015). The degree of clustering also depends on the radiation budget (Jakub and Mayer, 2017) and therefore plays a role in the cloud feedback. For an improvement of the representation of small scale convection in large scale models, it might therefore be beneficial to include a measure for the organization in a cloud population. A more detailed introduction on cloud sizes, spacing in cloud fields and convective organization is given in Chapters 3, 4 and 5 respectively.

1.4.2 Surface heterogeneity

Although clouds have been observed to organize in the absence of surface heterogeneities (Bretherton et al., 2005), heterogeneitiy might play an additional role. The origin of the formation of ShCu lies at the earth's surface. Convection transports moisture to higher altitudes where it can condensate and form clouds. Local surface heterogeneities alter the conditions for cloud formation and could therefore greatly affect ShCu. Moreover, surface induced flow patterns could impact the spatial distribution of clouds. If the amplitude of the heterogeneity is large enough, secondary circulations can develop (van Heerwaarden and de Arellano, 2008). The secondary circulations create areas where it is more favourable for clouds to form. In other words, in situations with secondary circulations clouds are not homogeneously distributed in space, and also their size might be influenced. A more detailed introduction on surface heterogeneity is given in Chapter 6.

1.5 Thesis objective

Studies have shown that clouds organize and that surface heterogeneity impacts the atmospheric flow. However, the impact of realistic surface conditions on convective organization is unknown, it is therefore the main objective of this thesis. The effect of the surface heterogeneity on cloud spatial patterns is quantified. For this, first some exploratory work is done on how to describe a cloud population in terms of size and spacing. The diurnal cycle of a cloud size distribution over land and the dependence of cloud spacing on cloud size are studied. Next, a comparison is made between different ways of quantifying organization in a cloud field. The results from these first studies are then used to answer the following question:

How does surface heterogeneity influence the spatial pattern of shallow cumulus clouds?

This thesis is divided in 7 chapters. First, in Chapter 2 more background information on shallow cumulus clouds and their representation in numerical models is given. This chapter includes a detailed outline of the thesis. Then follow 4 research chapters, with each their own detailed introduction. Chapter 7 ends with conclusions and an outlook.

Chapter 2

Theoretical background

2.1 Shallow cumulus formation

The formation of ShCu depends on the temperature and moisture conditions of the boundary layer as well as the surface conditions. In this section first the general structure of the boundary layer is described, followed by a description of ShCu formation over land and over the ocean.

2.1.1 Structure of the boundary layer

The conditions of the boundary layer over land are strongly coupled to a diurnal cycle. In the early morning hours the boundary layer is very shallow. This changes as soon as the solar energy starts to heat the surface. The surface absorbs the radiation and heats up, which increases the gradient of temperature between surface and atmosphere. As a reaction, the surface emits the absorbed energy. This is done either as sensible or as latent heat, the partitioning of the total energy in these two fluxes depends on the moisture availability.

The idealized structure of the boundary layer for potential temperature (θ) and moisture (q) is shown in Figure 2.1. The potential temperature is the absolute temperature, but corrected for adiabatic cooling or heating. Therefore, even if a parcel of air is cooled or heated, its potential temperature stays constant. This is visible in the profile of θ in a cloud free boundary layer (Fig. 2.1a). The area averaged θ is constant with height for the whole boundary layer, up to the inversion or boundary layer height (zi). The convection originating from the surface in the form of sensible and latent heat fluxes causes the air to mix, hence the constant θ and the name mixed-layer. Also q is in this layer well mixed and constant with height (Fig. 2.1c).

The increase in temperature close to the surface causes a parcel in this region to be less dense than its environment. It therefore has a positive buoyancy. The parcel is in an unstable environment and starts to rise. If the parcel does not experience any phase changes and does not mix with its environment, its temperature changes follow the dry adiabat. That means that the temperature of the parcel is decreasing with height. In the shallow morning boundary layer, the parcel won't condensate and will reach the top of the boundary layer without a phase change. At the top of the boundary layer it reaches the inversion. At the inversion are strong gradients in temperature and moisture. Here the parcel loses its positive buoyancy, but it still has some momentum from the upward motion. This momentum causes it to penetrate the inversion by a bit, thereby mixing some

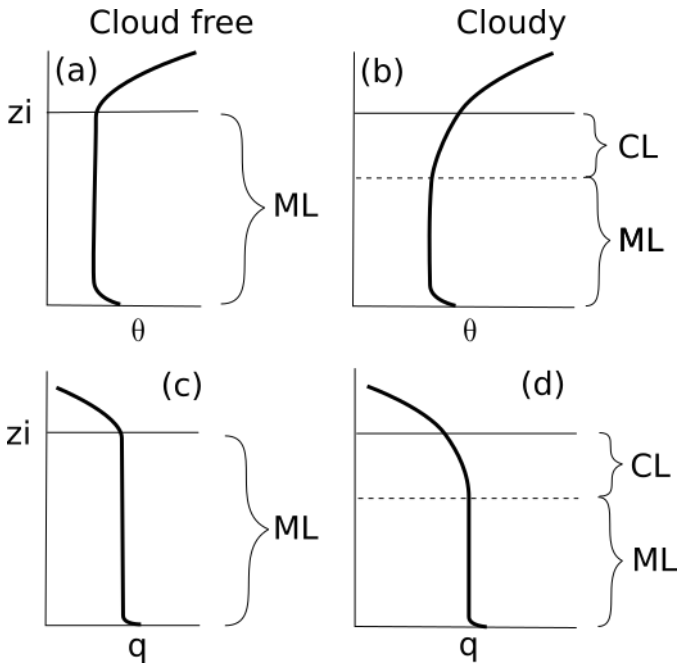

Figure 2.1: Idealized profiles for a cloud free (a,c) and cloudy (b,d) boundary layer. Shown are the profiles for potential temperature (θ) and humidity (q). Indicated are the boundary layer height (zi), the mixed-layer (ML) and the cloud layer (CL).

air from the free troposphere into the boundary layer. This process is called entrainment and it makes the boundary layer grow over time.

2.1.2 Shallow cumulus formation over land

The growth of the boundary layer results in rising parcels with enough inertia to reach their saturation point. The saturation point of a parcel depends on pressure and temperature, and it makes the water vapour of the parcel condensate. The presence of cloud condensation nuclei (CCN) enhances this process. The height at which condensation happens is called the lifting condensation level (LCL) and it marks the base of the clouds. The phase change of condensation above this height induces heat release and an increase of temperature. The increased temperature gives a parcel positive buoyancy again, causing it to rise further until it reaches the inversion height.

For a boundary layer with clouds the profiles of θ and q look slightly different than for a boundary layer without clouds (Fig. 2.1b,d). The water vapor condensation above the LCL causes latent heat release, this is visible in the increase of θ in the cloud layer. Since water vapor condensates, q decreases in this layer. Only a few parcels have enough inertia to reach the LCL and continue to rise and mix air. In the areas without clouds there is no convection and no mixing of air. Therefore a distinction is made between the mixing layer and the cloud layer. Because in both layers turbulence is present, here the combination of the two layers is defined as the boundary layer.

2.1.3 Shallow cumulus formation in the subtropics

Shallow cumulus not only form over land in the mid-latitudes, they also often occur over the ocean in the subtropics. The thermodynamic processes are similar for both locations,

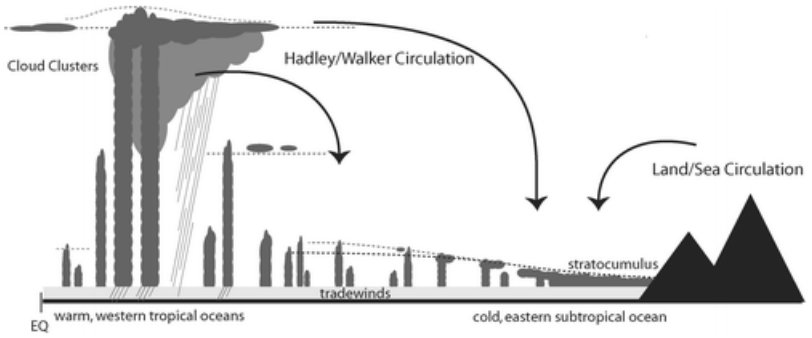

Figure 2.2: A schematic representation of the Hadley cell (Stevens, 2005)

but in the subtropics the main driver of convection is the typical circulation of air instead of the diurnal cycle of the heating of the surface. The steady large scale circulation is referred to as the Hadley circulation or Hadley cell (Tiedtke et al., 1988). A schematic is given in Figure 2.2. At the equator, ocean temperatures are high and winds converge, this causes the air to rise and creates favourable conditions for the formation of deep convection. Higher aloft the winds diverge north- and southwards, creating large areas of significant subsidence at higher altitudes. In these areas the dominant type of clouds are stratocumulus. The winds flowing back from these subsidence regions towards the equator are called the trade winds. Their name comes from the fact that they are steady and predictable and traders profited from them. These trade winds enhance evaporation of ocean water and trigger moist convection. The boundary layer therefore deepens which causes a decoupling of stratocumulus from the surface. The stratocumulus breaks open, this is called the stratocumulus to cumulus transition. The moist convection as a consequence of the trade winds results in the formation of Cu, in this region also called Trade wind cumulus. ShCu moisten the boundary layer, and this moisture is advected by the trade winds towards the equator where it promotes the formation of deep convection.

2.2 Observing shallow cumulus clouds

Because of their frequent occurrence, impact on the atmosphere and uncertain role in climate change, the study of shallow cumulus clouds is of importance. Observations of ShCu are very useful for this. However, observing cloud populations is not easy because of their complex character and large horizontal spread. Satellite measurements are a type of measurements able to observe entire cloud populations. Several satellite products are freely available, but they differ in resolution and observed locations. For cloud populations, satellite data has been used to study their spatial distribution (Joseph and Cahalan, 1990), also in comparison with model data (Tobin et al., 2012). Another source of observational data comes from point measurements. A good example of a measurement site that specifically focuses on clouds is JOYCE (Jülich ObservatorY for Cloud Evolution), located in south-western Germany (Löhnert et al., 2015). The aim of the observatory is to improve the understanding of cloud formation and precipitation processes. Measured are vertical profiles of temperature and humidity and cloud properties. One instrument that is currently being developed further is a scanning radar. This instrument has potential in providing 3D data about clouds and observing cloud size distributions in nature (Borque

et al., 2014). Further efforts in observing cloud populations include field campaigns where air planes fly through clouds and measure microphysical properties (e.g. Rauber et al. (2007)).

2.3 Modelling shallow cumulus clouds

Although input from measurements is absolutely necessary for our understanding of cloud processes, they cannot be used for weather and climate predictions and for sensitivity studies. For that numerical models have to be used. As described earlier, the representation of ShCu in said models is an important source of uncertainty. The main reason for this is their scale. ShCu are typically small and live shortly, in contrast to the large spatial and temporal scales on which numerical weather prediction (NWP) and climate models operate. The large scales are the only ones directly resolved by the model equations. Processes that are smaller than the grid spacing are called subgrid scale and unresolved, their contribution to the mean state is approximated by a parametrization scheme. Each parametrization scheme has its own uncertainties which contribute to the large uncertainty associated with cloud representation in climate models. Disentangling which effect comes from which scheme is difficult, the interaction between the schemes makes that the errors of one scheme might be compensated for by an other (Siebesma2004). An example of an uncertainty is present in the surface scheme. Here the Monin Obukhov Similarity Theory (MOST) is used to determine the surface heat fluxes, but the theory is only valid for homogeneous surfaces. In the case of a heterogeneous surface the total effect is approximated. The possible extra effects of surface heterogeneities are disregarded.

2.3.1 Parametrization schemes for shallow cumulus

The subgrid-scale ShCu are represented by a cloud scheme, which strongly interacts with the schemes for boundary layer processes, land surface, microphysics, radiation and convection. The cloud scheme determines the cloud cover from the thermodynamic conditions of the atmosphere, which are given by the convection and boundary layer scheme. It also takes care of the condensation and evaporation processes. The boundary layer scheme is responsible for the turbulent transport of heat, moisture and momentum which depend on the surface conditions given by the land surface scheme. All processes concerning the formation of cloud droplets and precipitation are included in the microphysics. The interaction with aerosols as CCN is present here as well. The radiation scheme determines the available energy in terms of radiation. Lastly, the convection scheme controls the transport of organized thermals.

These organized thermals in the convection scheme are problematic because they cover a large range of scales. They are present in large synoptic scale events, high and low pressure systems, fronts, thunderstorms as well as small individual clouds. Because of the large range of scales, convection is partly resolved and partly parametrized. To parametrize convection many options are available. Some schemes explicitly divide deep and shallow convection, while others opt for a unified approach (Hohenegger and Bretherton, 2011). Apart from handling the division of scales differently, other flavours are possible as well; one can e.g. include stochastics (Plant and Craig, 2008) or conditional Markov chains (Dorrestijn et al., 2013).

2.3.2 Grey zone of convection

The uncertainty associated with parametrized clouds can in theory be solved by increasing the model resolution. There has been much effort in this direction, supported by the ongoing development of technology and computer systems, making it possible to run atmospheric models on higher resolutions without having to compromise on domain size. However, the increased resolutions are not quite high enough to resolve all convection. At the same time, the common assumption in convection schemes that a thermal covers only a small part of a grid cell, is not valid any more at high resolutions. The in-between state of not resolving but also not accurately parametrizing convection is termed the grey zone, or Terra Incognita (Wyngaard, 2004). As a reaction to this problem, work is being done on scale-aware and scale-adaptive convection schemes. Depending on the grid resolution, the scale below which convection is parametrized is determined. One example of a scale-aware scheme is the ED(MF)^n scheme (Neggers, 2015). This scheme uses several plumes (updrafts), each representing a different scale. The foundation for a scheme like this has been laid by Arakawa and Schubert (1974), who introduced a cloud size distribution (CSD) based scheme. By using a scheme with a CSD at its foundation, the great variation of cloud sizes can be taken into account and at the same time their scale can be accounted for. For this approach to be successful, information is needed on the dependence of convective processes on cloud size and scale and an accurate description of the cloud population in terms of size is necessary. Studies suggest that the CSD can best be described by a power law with a scale break (Heus and Seifert, 2013). For scales larger than the scale break, the power law is not suitable any more. However, the reason for the scale break and its position are not well understood. Because of the importance of a correct representation of the shape of the CSD for the development of scale-aware convection schemes, the CSD is studied in more detail in Chapter 3 of this thesis.

2.3.3 Large Eddy Simulation

To study ShCu populations and their size distributions, we will not use observations or larger scale models with parametrized clouds, but a Large Eddy Simulation (LES) model. The advantage of a LES model is that it explicitly resolves the dominant scales of turbulent motion in the boundary layer, meaning neither clouds nor convection are parametrized and only the small subgrid scales are parametrized (Smagorinsky1963, Deardorff1970). Because of the high resolutions, it is computationally too expensive to do global simulations with an LES, but the regional scale domains can be large enough to study local cloud processes. An LES model provides 3D output with a high temporal frequency, which is necessary for studying cloud populations. LES models can be highly idealized in terms of large-scale flow or surface conditions, but they are a valuable tool to study isolated processes or the interaction between specific processes. By initializing LES for different regimes, many processes can be studied and the impact of different situations can be assessed with the help of sensitivity studies. Results from studies like this are beneficial for the formulation of parametrization schemes. In this study two different LES models with different set-ups are being used: DALES (Dutch Atmosphere Large Eddy Simulation, Heus et al. (2010)) and ICON-LEM (ICOsahydral Nonhydrostatic Large Eddy Model, (Zängl et al., 2014; Dipankar et al., 2015; Heinze et al., 2017)). The models and their set-ups are described in more detail in the chapters where they are used.

2.4 Thesis outline

The aim of this thesis is assessing the influence of surface heterogeneity on spatial patterns in shallow cumulus cloud populations. To this end, research first focuses on describing shallow cumulus populations in terms of size and spacing. In Chapter 3 we look into the cloud size distribution of shallow cumulus clouds. Given the strong diurnal cycle of ShCu formation over land, we employ many LES simulations for days with ShCu where for each simulation the daily cycle is captured. For the shape of the CSD a power law-exponential function is proposed, instead of a power law with scale break. The power law-exponential function captures the CSD over the complete range of cloud sizes and a scale break does not have to be taken into account.

Chapter 4 focuses on the relation between cloud size and cloud spacing. This relation is shown to be linear (Joseph and Cahalan, 1990). These findings were based on satellite observations, but nowadays models are available that are able to resolve ShCu over large domains. We employ an ICON simulation over the subtropical Atlantic which features many clouds per snapshot. With this data the dependence of cloud spacing on cloud size is studied. The cloud size dependency is a necessary piece of information for the development of CSD based parametrization schemes (Neggers et al., 2019). The results show that different definitions of cloud spacing result in different functional relations with cloud size. The spacing between clouds of a similar size is of special interest for scale-adaptive convection schemes, it is found to exponentially depend on cloud size.

In Chapter 5 several parameters to quantify convective organization are compared. The organization parameters are all applied in literature, but they all have their own advantages and disadvantages. For the comparison again the ICON data over the subtropical Atlantic is used, since it has many clouds and some interesting patterns in the cloud field. Not all compared parameters are able to capture the transition from more organized to less organized cloud fields, something that can be observed by eye from snapshots. Based on the results one parameter that performs best is selected for later use.

The results from previous chapters give the tools and knowledge to tackle the main objective presented in Chapter 6, which is the influence of surface heterogeneity on the shallow cumulus cloud populations. By using a set of ICON simulations with differing surface conditions, the impact of these surface conditions on the CSD and cloud organization can be assessed. The simulations are centred around JOYCE and differ in topography and land use type distribution. The sensitivity for these boundary conditions on ShCu is assessed for two different days. Small differences for the CSD are found, the slope is not affected by the surface conditions. However, the range of cloud size does change. For enhanced topography the maximum cloud size increases. A clear difference in organization can not be observed when looking at the organization parameter. Visual inspection and an extra analysis on the variance of vertical wind speed suggests that there might be differences between the fields that are not being picked up by the organization parameter.

Chapter 3

Investigating the diurnal evolution of the cloud size distribution of continental cumulus convection using multi-day LES

This chapter is published as: Thirza W. van Laar, Vera Schemann and Roel A.J. Neggers (2019), Investigating the Diurnal Evolution of the Cloud Size Distribution of Continental Cumulus Convection Using Multiday LES. *Journal of the Atmospheric Sciences*, vol. 76, no.3, doi: 10.1175/JAS-D-18-0084.1.

Abstract

The diurnal dependence of cumulus cloud size distributions over land is investigated by means of an ensemble of large-eddy simulations. 146 days of transient continental shallow cumulus are selected and simulated, reflecting a low mid-day maximum of total cloud cover, weak synoptic forcing and the absence of strong surface precipitation. The LES simulations are semi-idealized, forced by large-scale model output but using an interactive surface. This multitude of cases covers a large parameter space of environmental conditions, which is necessary for identifying any diurnal dependencies in cloud size distributions. A power law-exponential function is found to describe the shape of the cloud size distributions for these days well, with the exponential component capturing the departure from power law scaling at the larger cloud sizes. To assess what controls the largest cloud size in the distribution, the correlation coefficients between the maximum cloud size and various candidate variables reflecting the boundary layer state are computed. The strongest correlation is found between total cloud cover and maximum cloud size. Studying the size density of cloud area revealed that larger clouds contribute most to a larger total cloud cover, and not the smaller ones. Besides cloud cover, cloud base and cloud top height are also found to weakly correlate with the maximum cloud size, suggesting that the classic idea of deeper boundary layers accommodating larger convective thermals still holds for shallow cumulus. Sensitivity tests reveal that the results are only minimally affected by the representation of microphysics and the output resolution.

3.1 Introduction

Shallow cumulus clouds play an important role in Earth's weather and climate system. Both the mean amplitude and the variability of cloud cover have a significant impact on the earth's radiation budget (Cahalan et al., 1994). This requires a correct representation of the spatial structure of a shallow cumulus cloud field in climate and weather prediction models, in order to adequately simulate the radiative fluxes.

A complicating factor is that shallow cumulus clouds need to be parametrized because of their small and highly variable temporal and spatial scale. Various recent studies have identified such cumulus parametrizations to be at the heart of problems in both numerical weather prediction and climate simulations (Bony and Dufresne, 2005; Neggers and Siebesma, 2013; Vial et al., 2016). An active area of research which addresses a part of this problem focuses on the feedback of shallow cumulus clouds to climate perturbations (Zhang et al., 2013; Brient et al., 2015; Dal Gesso et al., 2015). Another problem for cumulus parameterization is the increasing resolution of the climate and NWP models and thereby the approach of the grey zone (Wyngaard, 2004). This means that parameterization schemes have to become sensitive for the resolution used in the model and be scale-adaptive. Moreover, it also means that our understanding of the spatial structure and diurnal cycles of cumulus cloud fields has to be improved. This is therefore actively researched at the moment (Arakawa and Wu, 2013; Dorrestijn et al., 2013; Kwon and Hong, 2016; Honnert, 2016).

In essence, making a cumulus scheme scale-aware means that size information somehow has to be included. This has recently motivated researchers to revisit the approach of formulating models in terms of cloud size distributions (CSDs), following Arakawa and Schubert (1974). Recent studies with CSD-based schemes by Wagner and Graf (2010), Park (2014), Neggers (2015) and Brast et al. (2018) report promising skill in reproducing scale-adaptivity, however closure of the CSD is still needed. This closure is still an open research question, and needs to be informed by reliable statistics on cloud sizes and their dependence on meteorological conditions.

Cloud size distributions for shallow convection have been investigated by numerous previous studies, using both observational and model data. Some older studies, like Plank (1969) and Wielicki and Welch (1986), report an exponential distribution, whereas later work commonly describes the functional form of the cloud size distribution with a power law:

$$N(l) = a\ l^b, \tag{3.1}$$

with l the cloud size and a and b fitting constants. A power law for cloud size distributions has been applied in observational studies like Benner and Curry (1998) and Zhao and Di Girolamo (2007) as well as LES studies like Neggers et al. (2003) and Heus and Seifert (2013). Jiang et al. (2008) used a power law as well to describe cloud size properties and they showed a good comparison between observations and model data. Recently, Feingold et al. (2017) used a power law distribution for their study on the relation between albedo and cloud cover. The power law exponent b from Eq. (3.1) describes the slope of the distribution. Typical values that have been found are between -1.7 and -2.5 (Rieck et al., 2014).

Many observational and modelling studies report the existence of a scale break in the power law (recently e.g. Trivej and Stevens (2010); Heus and Seifert (2013)), meaning that the pure power law fit is only valid for the small cloud sizes. For clouds larger than the scale

break the slope of the power law needs to be adjusted to fit the data, resulting in a double power law. How to determine the location of this scale break in a methodical way, as well as the underlying cause for this break in scaling, are still actively debated in the literature. Benner and Curry (1998) mention a break in fractal dimension at a location similar to the scale break in their CSD for which the underlying reason could be a maximum size of individual convective cells (Joseph and Cahalan, 1990), while Wood and Field (2011) hypothesise that the Rossby radius controls the characteristic cloud sizes and thereby the scale break. A third possible cause for the scale break is insufficient statistics; the sampling size is too small to capture the full distribution, in particular at the largest sizes which occur least frequently. Also domain size and resolution might influence the location of the scale break in a simulated cloud field.

Since it is hard to estimate the position of the scale break in a standardized way and its background is unclear, we follow Windmiller (2017) and Peters et al. (2009) in their approach of applying a power law-exponential function to the CSD. The power law-exponential fit is flexible to capture many shapes, from a pure power law to a more curved fit. By applying this fit we avoid having to explicitly deal with the scale break. The power law-exponential function is based on percolation theory and is defined as (Ding et al., 2014):

$$N(l) = a\ l^b\ exp(c\ l), \tag{3.2}$$

with c representing another fitting constant. Both the b and c are expected to be smaller than or equal to 0 and together they reflect the shape of the distribution. With decreasing c the distribution follows less a power law and is increasingly dominated by the exponential, especially at the large cloud sizes.

While the shape is one defining aspect of the CSD, its range in terms of size is another. An early study by Joseph and Cahalan (1990) suggests that the maximum cloud size scales with the depth of the boundary layer. More recently, Rieck et al. (2014) confirmed this dependence, but also reported that surface heterogeneity plays a role. Dawe and Austin (2012) found a strong relationship between the cloud area and the eventual height reached by the cloud. Sakradzija and Hohenegger (2017) showed that the shape of the distribution of cloud-base mass flux, which is closely related to the cloud size distribution, is determined by the Bowen ratio at the surface. While these new insights are encouraging, many of these studies were limited to single cases, or even single snapshots. Such snapshots do not contain any information on the diurnal evolution of the maximum cumulus cloud size. In addition, the use of single cases limits the statistical significance of the obtained results. To assess the robustness of the results and to obtain information on the diurnal signal of the maximum cloud size one would need a database of many different cases covering a broad parameter space of large-scale conditions.

For this study, a library with 146 individual cumulus days is created to investigate the behaviour of the size distribution during diurnal cycles of cumulus over land. Continental shallow cumulus cases are chosen because of their transient nature, with a strong temporal evolution in the amount of clouds, their elevation and depth, and the size distribution. Using many different days over the course of 5 years enables us to do statistical studies on cloud size controlling factors on a diurnal time-scale. For all the selected days a 24-hour Large-Eddy Simulation (LES) simulation is performed, and a clustering algorithm is applied to compute the cloud size distribution and study both its shape and the maximum cloud size. Correlations between the maximum cloud size and a set of candidate variables are calculated to establish the role of these variables in controlling the upper limit of the distribution of cloud sizes. Sensitivity tests are performed with a subset of cases on the output resolution and the microphysics representation.

The configuration of the LES experiments and the case selection is described in section 2. In section 3 the clustering algorithm and the derivation of the cloud size statistics is explained, followed by the results in section 4 and 5. Finally, discussion and conclusions can be found in sections 6 and 7.

3.2 Multi-day LES

The derivation of statistically significant size distributions requires 3D cloud fields that contain a sufficient number of clouds and are available at a frequency high enough to resolve any diurnal signal. While the latest scanning radar strategies show promising capability in detecting CSDs in nature (Borque et al., 2014), this technique is still in its infancy and needs to be fully explored. We therefore still rely on LES simulations to complement observational data with 3D fields at high frequencies (Neggers et al., 2012; Schalkwijk et al., 2015; Zhang et al., 2017). For this study daily LES as applied at the meterological supersite JOYCE (Jülich ObservatorY for Cloud Evolution) in Germany will be used.

3.2.1 JOYCE

The observational supersite JOYCE (Löhnert et al., 2015) is a continental mid-latitude site, well suited to study diurnal cycles. It is equipped with state of the art cloud detection instrumentation which are operational on the long-term. Mid-latitude cumulus are regularly present at JOYCE during summer, since the site is close enough to the sea to ensure that low-level humidity is frequently high enough to allow daytime cumulus formation. The abundance of long-term observations allows for confronting LES with relevant observational data, this is however not the focus of this study and is considered a future research topic. The measurements taken at JOYCE will only be used for the selection of shallow cumulus days and a short comparison of the measured and modelled cloud cover.

3.2.2 Cumulus day selection

Since the focus of this research is on boundary layer cumulus clouds, days which show this cloud regime have been selected to be simulated. Several specific criteria are applied to visualized reflectivity of lidar and radar observations taken at JOYCE. Selected days do not show any precipitation and no significant synoptic scale activity during daytime. They do have a significant period of cumulus convection around noon and an increase of the LCL during the day. Cloud covers are small and the cumulus events are isolated, which means that they are not influenced by synoptic events at the beginning and ending of the day. These criteria are applied on the summer months of the years 2012-2016, in the winter months conditions are usually too cold for shallow cumulus to form. The result is a selection of 146 days in total. How these days are distributed over the years and the months can be seen in Figure 3.1. The resulting library of cases provides us with a parameter space of considerable width. By simulating all 146 days individually, and not simulating a single composite like Zhang et al. (2017), the internal variability of the ensemble is resolved. This provides additional information on the robustness of any diurnal signal in the CSD properties, as well as its spread for the environmental conditions covered by the library of cases. Since we are interested in the diurnal cycle, only the daytime hours between 9 and 17 hr are used for further analysis.

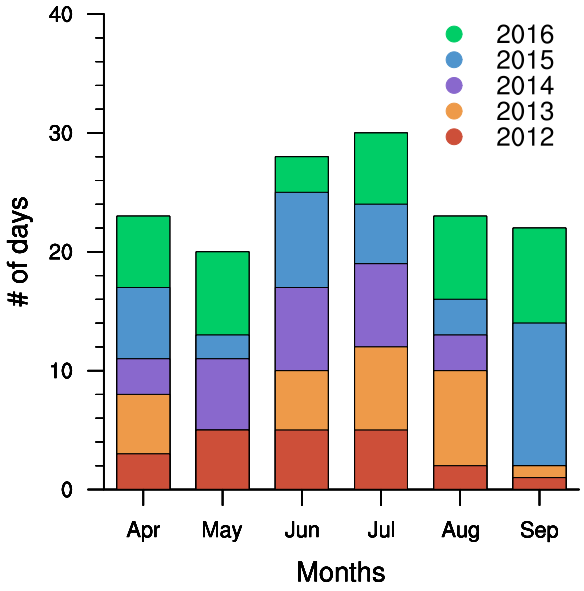

Figure 3.1: Histogram showing how the total of 146 days are spread over the months and years.

3.2.3 DALES

The Dutch Atmospheric Large-Eddy Simulation (DALES, Heus et al. (2010)) model is used for the simulation of the selected days. DALES has taken part in many recent intercomparison studies on shallow cumulus convection (e.g. van Zanten et al. (2011); van der Dussen et al. (2013)) which document its skill in simulating this cloud regime. The land surface parametrization scheme is described in Heus et al. (2010) and it is similar to the ECMWF surface scheme. In the simulations, the state of the land surface is evolving with time-varying temperature and soil moisture. As a result, the surface energy fluxes are also time-varying, calculated interactively through the surface energy budget. For the calculation of the vertical transfer of radiative energy Monte Carlo Spectral Integration is used, as described in detail by Heus et al. (2010). The resolved turbulent domain reaches up to 5 km. Above this height, the vertical profile of the thermodynamic and cloudy state of the IFS analysis is used to calculate the downward transfer of radiative energy into the turbulent domain, by which the impact of clouds overhead are accounted for. In addition, a climatological vertical profile of ozone is provided to the radiation scheme. The domain size of the simulations is 12.8x12.8x5 km, with a resolution of 50 m in the horizontal and 40 m in the vertical direction. Simulations start at 00:00 LT (LT leads UTC by two hours in European Summer in Germany) and cover one full day (24 hours). Every 15 minutes, every 4th point of the three-dimensional field of all required model variables is saved for further analysis. This sub-sampling is done in order to keep the amount of data manageable. The impact of the reduced data output on the CSD will be assessed (as reported in the Appendix, section 3.8), showing only marginal dependence, which supports taking this approach. Zhang et al. (2017) report a dependence of the total cloud cover on the featured micro-physics scheme. In the simulations the two-moment parameterization scheme of Seifert and Beheng (2004) is used, but its dependence on the results is tested. A sensitivity study is performed by using some additional simulations with a simple non-precipitating all-or-nothing cloud scheme by Sommeria (1976) that are done for the years 2013 and 2014.

3.2.4 Initialization, boundary conditions and large-scale forcing

The 24-hour long simulations are initialized and driven by time-varying boundary conditions and large-scale forcings derived from the Integrated Forecasting System (IFS) of the European Centre for Medium-range Weather Forecasts (ECMWF). The method as applied here to drive the LES model was first described by Neggers et al. (2012), and was recently used in slightly modified form in the study by Gesso and Neggers (2018). All fields provided to the LES as input are based on two analysis fields per day, at 00:00 and 12:00 UTC, supplemented by short-range forecast fields to cover the intermediate timepoints at 3, 6 and 9 hours. This effectively yields a forcing dataset covering 24 hours at 3 hour time-resolution, which is assumed sufficient to resolve diurnal and synoptic signals in the large-scale forcing.

The initialization of the height-dependent atmospheric state variables concerns the wind components u and v, the liquid water potential temperature θ_l, total water specific humidity q_t, the cloud state variables liquid and ice condensate q_l and q_i, and ozone (for the radiation scheme). Also initialized are the soil temperature and moisture, the latter scaled to preserve the fraction between the field capacity and wilting point (Neggers et al., 2012). After initialization, all these variables are allowed to evolve freely, apart from ozone which is not changing with time.

The prescribed boundary conditions at the surface include the surface geopotential, the albedo, the roughness lengths for heat and momentum, the leaf area index (LAI), and the near-surface air pressure. For both the LAI and the roughness lengths the IFS values at this location are extracted from the MARS archive (ECMWF, 2017). Only the near-surface air pressure changes with time, determining the geometric height of the pressure levels in the profile above. The soil temperature and moisture can evolve freely but interact with the atmosphere, through the surface energy and humidity budgets. In this respect the LES experiments in this study differ from those in other recent LES studies at supersites, which typically use a fully prescribed soil state (Heinze et al., 2016; Gustafson et al., 2017).

The impact of large-scale weather systems, which can be significant at mid-latitude locations, is represented by means of prescribed advective forcing tendencies in the budget equations of the four atmospheric state variables $\{q_t, \theta_l, u, v\}$. This forcing is both height- and time-dependent, and is also continuous, being interpolated linearly between the 3-hourly IFS time-points. IFS data at pressure levels is used to this purpose, in order to avoid problems with advective calculations in areas of steep orography that arise when terrain-following coordinates are used (Simmons and Burridge, 1981; Simmons, 1986). Orography is not included in the simulation, justified by the fact that the direct vicinity of the site is relatively flat. The horizontal tendencies are calculated from the wind components and the local gradients on the 0.1x0.1 degree operational IFS grid, averaged over a 0.5x0.5 degree horizontal area around the site of interest to obtain smoothly varying forcing fields. The vertical component of large-scale advection is calculated interactively from a prescribed subsidence profile that acts on the vertical gradients in the model state variables. Momentum is forced using a prescribed large-scale horizontal pressure gradient, appearing in the resolved Navier-Stokes equations as a prescribed geostrophic wind, and by accounting for the Coriolis force. The momentum equations take the form as discussed in any standard academic textbook on atmospheric dynamics, for example Holton and Hakim (2012).

Finally, following Neggers et al. (2012) the profiles of $\{q_t, \theta_l, u, v\}$ are continuously nudged to the IFS-derived state at a synoptic time-scale of 6 hours. Relaxation at this synoptic time-scale has been proven tight enough to effectively limit excessive model drift, but still loose enough to allow the turbulence to act freely.

Figure 3.2 gives some insight into the typical summertime large-scale forcing at the JOYCE

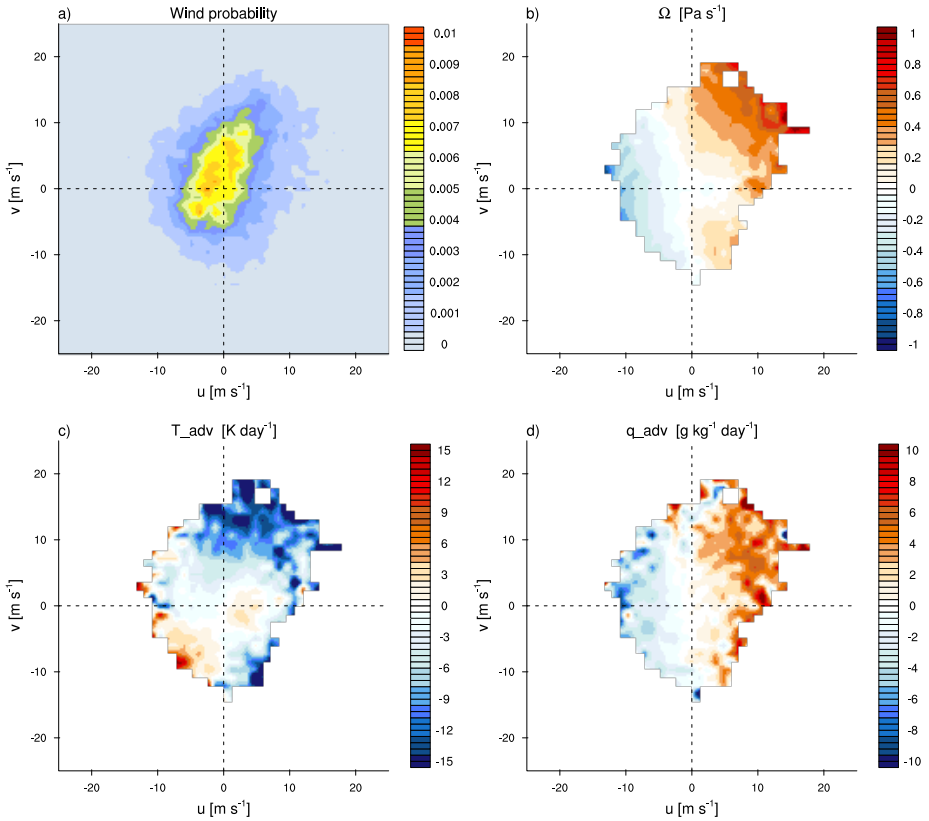

Figure 3.2: A five-year climatology of large-scale forcing at 500 m height at the JOYCE site for the months April-September in the period 2012-2016, as derived from IFS data. Panel a) is the probability density function of wind speed, while the other three panels are time-averages of b) pressure velocity Ω, c) temperature advection and d) humidity advection.

site, derived from the IFS fields for the period 2012-2016. The focus is on a height of 500 m that is normally inside the daytime boundary layer. Some features are visible that can be considered typical for midlatitude near-coastal sites in Northwestern Europe, making it different from for example the ARM SGP site (Zhang et al., 2017). First, stronger wind speeds are predominantly from the south-southwest, a clear fingerprint of the polar jetstream and its quasi-geostrophic forcing in this area during summer. Second, there is a rough east-west divide in the sign of the temperature and humidity advection (Fig. 3.2c,d), with westerlies associated with moistening and cooling and easterlies with warming and drying. This is explained by the proximity of the ocean in the west, which tends to cool and moisten the coastal regions during summertime. In general, the amplitude of the advection increases with wind speed, as can be expected from its definition. Interestingly, the large-scale subsidence roughly correlates with the advective tendencies, with strong westerlies being associated with stronger subsidence, and easterlies with lifting. The latter could partially reflect the cases of convective instability in the summertime associated with low surface pressure and strong convergence.

3.3 Methods

3D instantaneous snapshots of liquid water and buoyancy from the LES serve as input for the derivation of the CSDs. Alternatively, one can track individual clouds through

their life-cycle and thereby calculate their size (Heus and Seifert, 2013). To use snapshots instead is motivated by two reasons. Firstly, the typical life-time of a single cumulus cloud is much shorter than the diurnal signal. Second, using snapshots significantly simplifies the analysis, which is essential for processing the 146 daily simulations with multi-dimensional high-frequency output.

3.3.1 Cloud definitions

The starting point is to define a single LES gridbox as either cloudy or non-cloudy. Two different criteria for cloudiness are used; 1) the liquid water content is greater than zero or 2) the liquid water content is greater than zero and the buoyancy is positive as well. The combined criteria of condensate presence and positive buoyancy is often referred to as the *cloud core* (Siebesma and Cuijpers, 1995). This option is added because it is suggested in literature that clouds with and without positive buoyancy have different dynamics and a different vertical extent and should therefore be considered separately (Zhang and Klein, 2013). Unless otherwise indicated, in this study only the liquid water content of a gridbox is used as criterium for cloud presence.

3.3.2 Clustering algorithm and CSD calculation

Details of the computation of the cloud sizes and their distribution can be found in Neggers et al. (2003), but the basics will be described here. All horizontally and vertically neighbouring cloudy grid boxes are clustered into clouds and their size is calculated. Cloud size is defined as the square root of the covered horizontal area, where this area is the number of cloudy grid cells times the area per grid cell. There are two ways to determine the horizontal area covered by the cloud, either as the projected area, or as the horizontal area averaged over the depth of the cloud (for mathematical definitions see Neggers et al. (2003)). By using the projected area it is assumed that cloud sizes do not vary much with height and that horizontal wind does not influence the aspect ratio. Cloud statistics based on the projected area are better suitable for comparison with satellite observations. Therefore the projected cloud cover will be the default definition in this study. All sizes of the population in a single snapshot are sorted into a histogram, which is a discretized form of a CSD. The maximum cloud size ($l_{\max}$) which is used later is defined as the biggest cloud of the field.

Because of the sub-sampling of the original LES output, the resolution of the model output fields is four times smaller than the simulation resolution of 50 m. The derived cloud sizes are therefore organized in bins of 200 m. The amount of clouds per bin is divided by the total number of clouds in the field, resulting in the normalized cloud size distribution (CSD, denoted as $N^*(l)$). The computed CSDs with a frequency of 15 minutes are averaged per hour. This yields a time-frequency at which the diurnal cycle can be considered well resolved, while at the same time the statistic reliability is increased.

3.4 Results: exploring the variability

The 146 LES simulations produce a large amount of data with considerable variability. First some average properties and individual days will be studied. The findings based on a subset of the data will guide a more thorough analysis using the complete dataset in the next section.

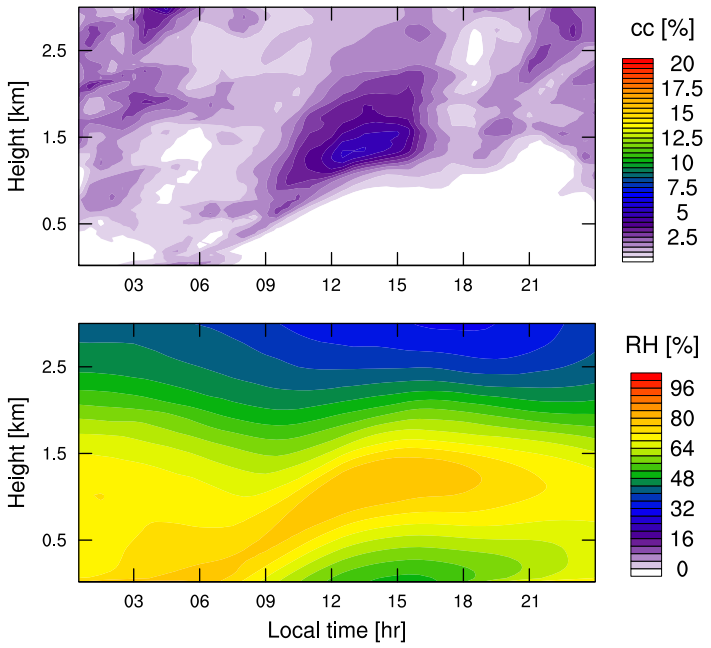

Figure 3.3: The simulated cloud cover (upper panel) and relative humidity (lower panel), averaged over all the simulated days.

3.4.1 Averaged mean state

In the average cloud fraction of all simulated days (Fig. 3.3a) some typical features in the vertical structure for diurnal cycles of shallow cumulus over land can be recognized and it shows many similarities with e.g. Zhang et al. (2017). First, during the day the well-defined cloud base rises in a quasi-linear way. Second, the layer with high cloud fractions deepens gradually, reaching a depth of about 500-1000 m. Third, the averaged cloud cover does not exceed 10%, which is typical of fair-weather cumulus (Nuijens et al., 2015). These results suggest that the ensemble of selected cases matches the prototype behaviour of continental shallow cumulus as reported in previous studies (Brown et al., 2002; Zhang et al., 2017).

However, there are some features visible in the figure that are different from previously mentioned studies, but typical for our specific case of mid-latitude continental shallow cumulus at a site in relative proximity to the coast. Due to morning fog there is a low cloud cover close to the ground before sunrise. With time this fog breaks up, resulting in low lifting condensation levels during the early morning. There is also a significant cloud cover during the night, this probably points at some synoptic activity.

Similar features as for cloud fraction can be seen for averaged relative humidity (Fig. 3.3b). There is a relatively high humidity in the cloud layer, starting low at the surface in the early morning hours because of fog. The cloud layer gradually grows deeper during the day. The lowest levels of humidity at the surface are found at the middle of the day. During these hours the moisture is effectively transported upwards, a prerequisite for cloud formation at greater heights. Besides this, drier air from the free troposphere is transported to the surface by effective entrainment at the inversion height (Betts, 2009).

3.4.2 Cloud cover comparison

Cloud cover is an important variable in this study. Therefore, despite a comparison of model and observational data not being a primary goal of this study, the modelled cloud

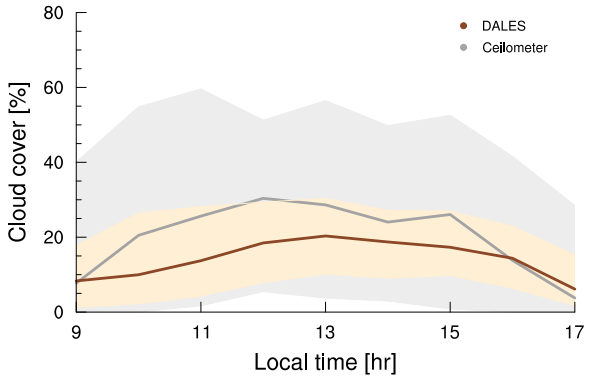

Figure 3.4: The diurnal evolution of the median (solid lines) and upper and lower quartiles (shading) of cloud cover as observed by a ceilometer (grey) and cloud cover as modelled by DALES (dark red).

cover will be shortly evaluated using ceilometer data. For all the 146 selected cumulus days the cloud cover up to 3 km is determined from the ceilometer observations and from the LES output. The median plus upper and lower quartiles of both ceilometer and simulated cloud cover as a function of time are shown in Figure 3.4. The spread in the ceilometer data is considerably larger than for the model data and the median cloud cover during the day (except for the last hour) is 5-10 % smaller for the model data. Both observations and simulations do however show the typical diurnal cycle for shallow cumulus of an increased cloud cover in the morning and decreased cloud cover in the afternoon.

3.4.3 Individual cumulus days

There is much variability among the days that establish the averaged behaviour that is shown in Figure 3.3. Five random individual days are selected to showcase this variability, each with a different cloud cover around noon (Fig. 3.5). The two days with the lowest cloud covers (28 May 2012 and 23 June 2016) both show a thin cloud layer around noon. The main difference between these two days is the development of the depth of the cloud layer. For 28 May 2012 this happens gradually, whereas for 23 June 2016 there is already a peak in cloud depth and cloud cover during the early morning. On 29 July 2012 the cloud layer stays shallow during the first few hours of the day, but around noon it suddenly deepens. 9 September 2015 shows a day with a low LCL in the morning, this might be because of the synoptic activity during the night. There are also high cloud fractions at the top of the thick cloud layer. This suggests the presence of a layer just below an inversion with some more stratiform, capping outflow. On 18 April 2015 there is an increasing cloud base during the day, a quite shallow cloud layer, but nevertheless high cloud fractions between 40 and 50 %.

3.4.4 Cloud size distributions

The considerable variation in the cloud structure visible in Figure 3.5 is also reflected in the shape and range of the associated CSDs taken at 13:00 LT (Fig. 3.6). The dots in the figure represent the discretized distribution and the lines in the corresponding colours the fit, following the power law-exponential function from Eq. (3.2). All details, including the constants of the fit and the maximum cloud size of the distributions are given in Table 3.1. Judging by eye, the power law-exponential function seems to fit the data well, since

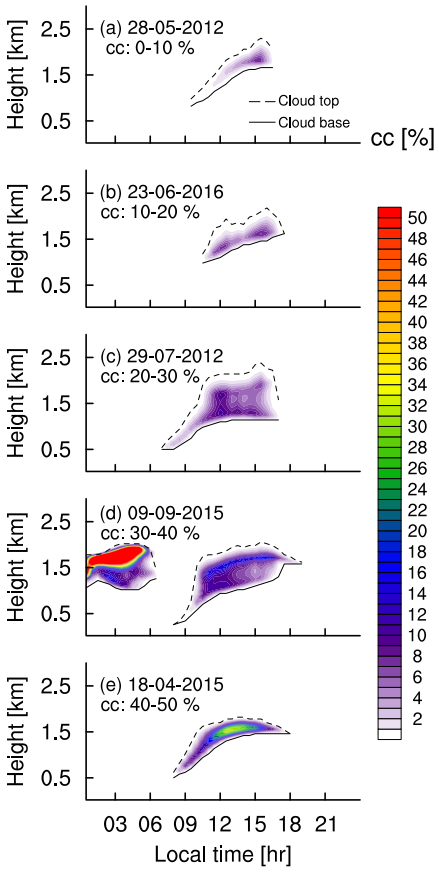

Figure 3.5: The cloud fraction for five simulated days, all with a different maximum cloud cover at 13:00 LT. The dashed line represents highest cloud top, the solid line lowest cloud base.

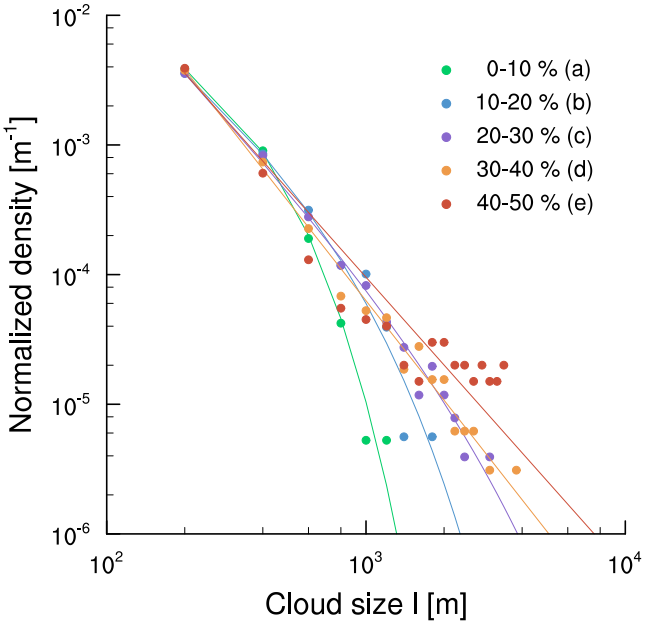

Figure 3.6: The cloud size distribution for 13:00 LT for five different days (see also Table 3.6). Each color represents a different day with a different cloud cover.

Table 3.1: Days shown in Figure 3.6 with their cloud cover, fitting constants, maximum cloud size and colour as used in the figure.

Panel Fig. 3.6	Day	cc [%]	a	b	c (x 1000)	$l_{\max}$	Colour in Fig. 3.6
a	28-05-2012	9	0.02	0.0	-7.4	1200	Green
b	23-06-2016	14	12.10	-1.45	-2.2	1800	Blue
c	29-07-2012	27	347.79	-2.15	-0.5	3000	Purple
d	09-09-2015	35	2809.85	-2.55	0.0	3800	Orange
e	18-04-2015	46	536.88	-2.25	0.0	3400	Red

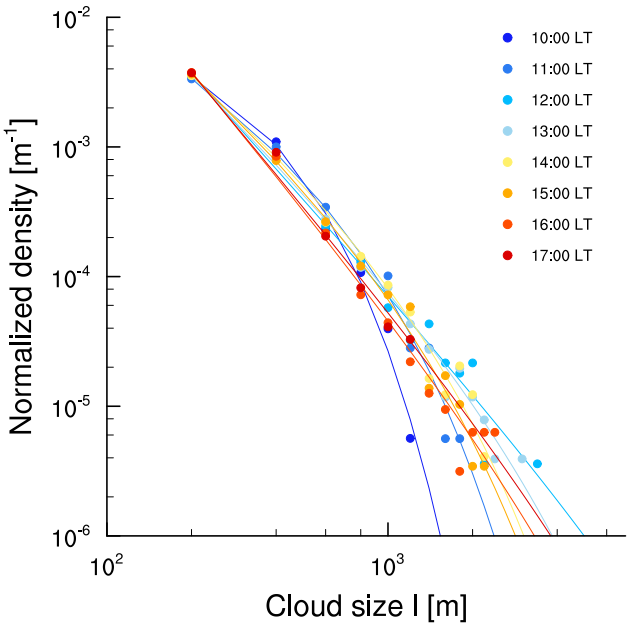

Figure 3.7: The cloud size distributions for all available hours of 29 July 2012 (see also Table 3.2).

for most days shown in the figure it captures the power law part, as well as the tail of the distribution. Three features stand out in these CSDs. Firstly, an increasing projected cloud cover largely coincides with an increasing simulated maximum cloud size ($l_{\max}$). Secondly, the power law scaling is visible for the smaller cloud sizes, for larger cloud covers it is also visible for the larger cloud sizes. The small values for c in Table 3.1 indicate that for larger cloud covers the power law dominates across all scales. Thirdly, the data shows more noise at larger cloud sizes, especially for larger cloud covers (shown in red). We speculate that the reason behind this is the undersampling of the least frequently occurring larger cloud sizes. This suggests that statistics need to be improved to obtain convergence. In this study the undersampling at larger sizes is addressed by considering many more days, since due to computational feasibility the domain size and sampling frequency are limited.

As an example of the diurnal evolution, the CSDs for 29 July 2012 (Fig. 3.5c) are shown in Figure 3.7. The values for b, c, cc and $l_{\max}$ can be found in Table 3.2. The diurnal cycle of $l_{\max}$ shows a peak at midday, a similar pattern is present for the cloud cover. The fitting constants b and c have a somewhat different pattern. Over the course of the day, the b decreases while the c increases.

Table 3.2: Fitting constants, cloud cover and $l_{\max}$ for figure 3.7 of the hourly CSDs of 29 July 2012.

Time of day	b	c (x 1000)	cc	$l_{\max}$
10	0	-6.1	10.3	1200
11	-1.25	-2.3	15.0	1800
12	-2.4	-0.1	29.2	3400
13	-2.15	-0.5	26.7	3000
14	-1.75	-1.2	23.5	2200
15	-1.9	-1.2	24.6	2200
16	-2.6	-0.3	23.0	2400
17	-2.55	-0.2	6.8	1200

3.5 Results: shape of the cloud size distribution

The CSDs for the individual days suggest a relation between cloud cover and $l_{\max}$, and between cloud cover and the shape of the distribution. Now the complete dataset with all its variability will be used to investigate these relations in more detail and increase the statistical reliability.

3.5.1 Evolution of the fitting constants

As a first step in a more detailed analysis the power law-exponential fit is determined for all days and times available. This results in a set of fitting constants a, b and c. Since the constants b and c reflect the shape of the fit, they are plotted against cc in Figure 3.8. The data is coloured by $l_{\max}$, revealing that $l_{\max}$ correlates well with cloud cover. The fitting constants b and c are decreasing and increasing with cloud cover respectively, with c becoming less negative and closer to zero. For larger cloud covers the b gets more negative, which means that the slope of the power law part of the distribution is steeper. What processes influence the slope of the power law is still a matter of debate. Sakradzija and Hohenegger (2017) proposed a mechanism based on the Bowen ratio, whereas Feingold et al. (2017) argued that the oscillation of the cloud field between a relative abundance of small or large clouds could explain the steepness of the slope.

The diurnal evolution of the b and c constants with the median (solid line) and upper and lower quartiles (shaded area) results from the fit of Eq. (3.2) to the hourly CSDs of all days (Fig. 3.9). In the early morning hours when the boundary layer is rapidly deepening and clouds start to form, the power law becomes more dominant (decrease of b, less negative c). During the middle of the day the b is low and the c is close to zero, whereas at the end of the day the c becomes more negative again and the exponential part of the fit becomes stronger again. Apparently, for this library of cases the shape of the CSD changes significantly during the day, with a shift from exponential at cloud onset to more power law at midday and back to exponential in the evening. Interpreting this result purely from the perspective of percolation theory would mean that a fit closer to a pure single power law implies an enhanced aggregation of cumulus clouds (Windmiller, 2017). However, this should be applied with caution; it does not have to mean that individual cumulus clouds actually merge during their ascent. All it reflects is that the cloud objects are perhaps bigger and are situated closer together, so that the system is closer to critical percolation.

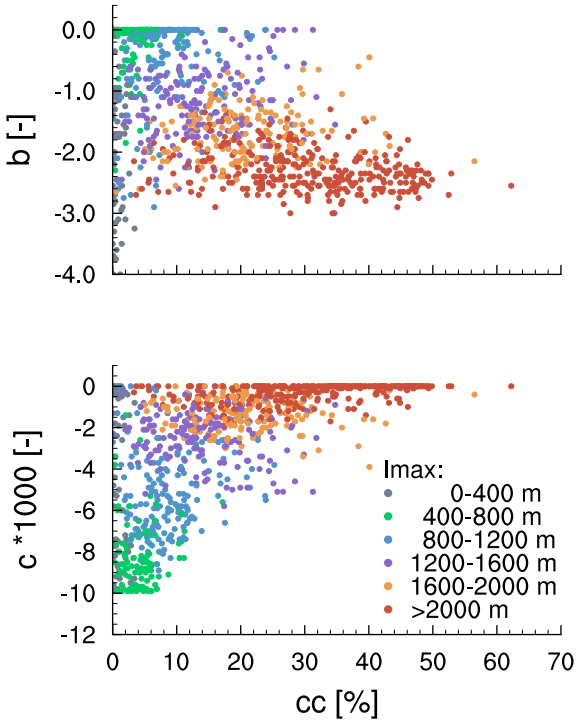

Figure 3.8: Cloud cover against fitting constants b and c associated with the power law-exponential fit from Eq. (3.2). Each dot indicates the properties of a single hourly CSD, with the color indicating the maximum cloud size.

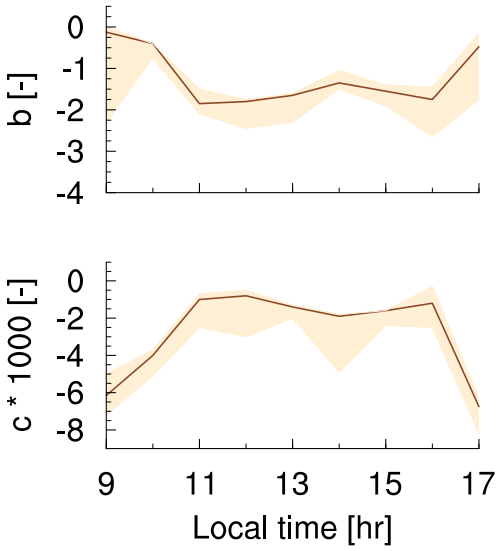

Figure 3.9: Temporal evolution of fitting constants b and c associated with the power law-exponential fit (Eq. 3.2) as applied to the hourly cloud size distribution.

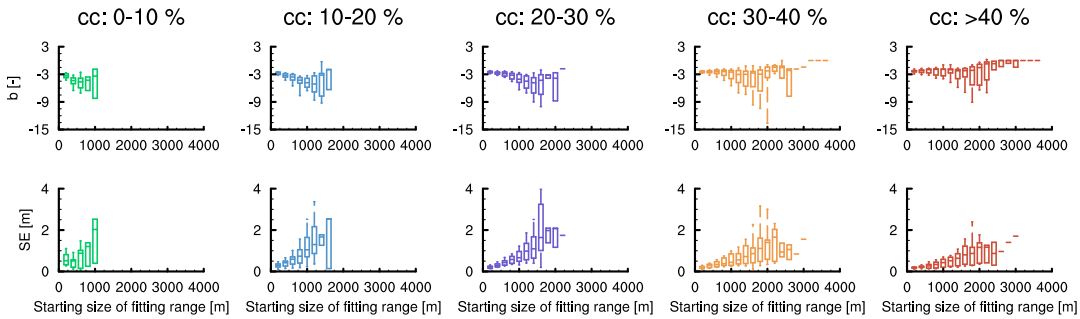

Figure 3.10: Box-whisker plots of the exponent of the power law (upper row) and the standard error of the exponent SE (lower row). The x-axes denote the starting size of the fitting range, so going to the right means less and less cloud sizes are taken into account. Every column presents the data for a different cloud cover bin.

3.5.2 Interpreting the power law fits

As discussed earlier, power law functions (Eq. (3.1)) have been commonly used in the past to fit to CSDs, in contrast to power law-exponentials. A major drawback of a single pure power law is that it only holds for the smaller cloud sizes and does not describe the distribution for the larger cloud sizes. The slope of the CSD can change considerably across the size-range, sometimes featuring abrupt changes. To gain more insight into the dependence of the CSD slope on cloud size and cloud cover, we take advantage of the number of simulations at our disposal.

The following 'goodness-of-fit' method is applied on all CSDs at 13:00 LT. First, for each day a single power law fit is applied to the full width of the CSD, covering all size-bins. This yields an exponent b and a standard error (standard deviation) of this b (SE). The fit is then applied on a reduced number of bins, leaving out the first bin at the lower end of the spectrum. The result is a new exponent and SE. This is repeated until the minimum amount of three bins to perform a fit is reached. The slope of the fit will vary with the amount of bins taken into account. This method highlights the variation of the CSD slopes across the size-range, giving the opportunity to assess its dependence on the cloud sizes that are covered.

The results are visualized in Figure 3.10. Box-whisker plots for every cloud cover bin show the median values and the spread in the data. The x-axes denote the cloud size of the starting bin for the fitting range. Going from left to right means using less and less cloud sizes for applying the power law fit, since the starting cloud size becomes bigger. For both the exponent (upper row) and the SE (lower row) the divergence increases towards a smaller fitting range.

The values of the exponent b at cloud size 200 m (smallest bin) compare well to literature (Rieck et al., 2014). The increasing divergence from this value depends on the cloud cover. This divergence can also be seen for SE, and can be explained in two ways. At least partially because of the statistics being less robust for a shorter fitting range, the uncertainty and therefore the SE becomes higher. However, it also shows that especially for smaller cloud covers the power law is not sufficient to properly describe the CSD. Despite the statistical noise, the stratification as a function of cloud cover shows that for higher cloud covers the variations in b and SE are smaller than for lower cloud covers. This confirms the results from the previous section, that at larger cloud covers the power law reflects the shape of the CSD on a broader range of sizes. Since the exponential part is not taken into account

in this analysis, the fit for low cloud covers is less good than for large cloud covers and there is more variation in the slope.

3.5.3 Maximum cloud size

Apart from its shape, the CSD derived from the simulations is also defined by the maximum cloud size ($l_{\max}$) that it covers. Several physical processes are candidates for controlling this $l_{\max}$. The boundary layer height for example, limits the vertical size of the largest convective element that can be accommodated in the layer, an argument often used in boundary layer scaling. This in turn could also constrain the horizontal size of the element, although it is already shown by de Roode et al. (2004) that this is not the case for stratocumulus. The surface heat fluxes control the boundary layer height, and thereby they also might control the maximum cloud size. Recently, Sakradzija and Hohenegger (2017) showed that the Bowen ratio controls the shape of the mass flux distribution through the efficiency of the moist heat cycle in the subcloud layer. The shape of the mass flux and the cloud size distribution is similar, so therefore the surface flux partitioning could also have an influence on the cloud size. Lastly, humidity of the environmental air at the top of the boundary layer might have an impact on the maximum cloud size. When the environmental air is more humid, clouds get less diluted at the edges and can grow bigger. Dry surroundings may limit the horizontal growth of the cloud by making cloud evaporation more efficient at the cloud edges. A priori it is not known which processes will control $l_{\max}$ since this has not been studied before. However, the considerable variation in cloud structure and amount, covered by the library of cases in this study, might provide a large enough parameter space to gain more insight.

As a first step, the ensemble-mean time evolution of several candidate variables expressing the state of the boundary layer are compared to that of $l_{\max}$ (Fig. 3.11). The solid line in each panel is the median over all cases, while the shaded area represents the lower and upper quartile. To study the relations between the variables and $l_{\max}$ the Pearson correlation coefficient is calculated between all individual data points as:

$$R_{x_1,x_2} = \frac{COV(x_1, x_2)}{\sigma_{x_1}\ \sigma_{x_2}}, \tag{3.3}$$

and given in the upper right corner of the panels in Figure 3.11. COV is the covariance and σ the standard deviation. The order of the panels reflects the degree of correlation with $l_{\max}$. The time series of both the convective velocity scale (w_*, panel a) and the buoyancy flux (panel d) show the typical evolution of a daytime convective boundary layer, with turbulence increasing during the morning and decaying again towards sunset. The Bowen ratio (panel b) shows on average a decrease during the whole day, explaining the poor correlation with $l_{\max}$. The relative humidity (RH, panel c) averaged over the subcloud layer shows a decrease as the boundary layer develops and becomes deeper. Minimum cloud base (zb, panel e) shows the ballistic growth that has been described earlier by Stevens (2007) for this type of boundary layer. Boundary layer depth (zi, based on the Richardson number, panel f) increases strongly during the early morning hours and after that continues to deepen at a lower rate. The convective inhibition (CIN, panel g), maximum cloud layer top height (zt, panel h), and the cloud cover up to 3 km (cc, panel i) all show a similar evolution. There is an increase in the beginning of the day, and a decrease at the end of the afternoon. This behaviour better resembles the evolution of $l_{\max}$. The cloud cover up to 3 km (cc, panel g) shows the strongest diurnal cycle of all candidate variables. The fact that CIN, zt, and cc have well-defined maxima during the afternoon explains their good correlation with $l_{\max}$ (panel j), which shares this feature. The evolution of $l_{\max}$

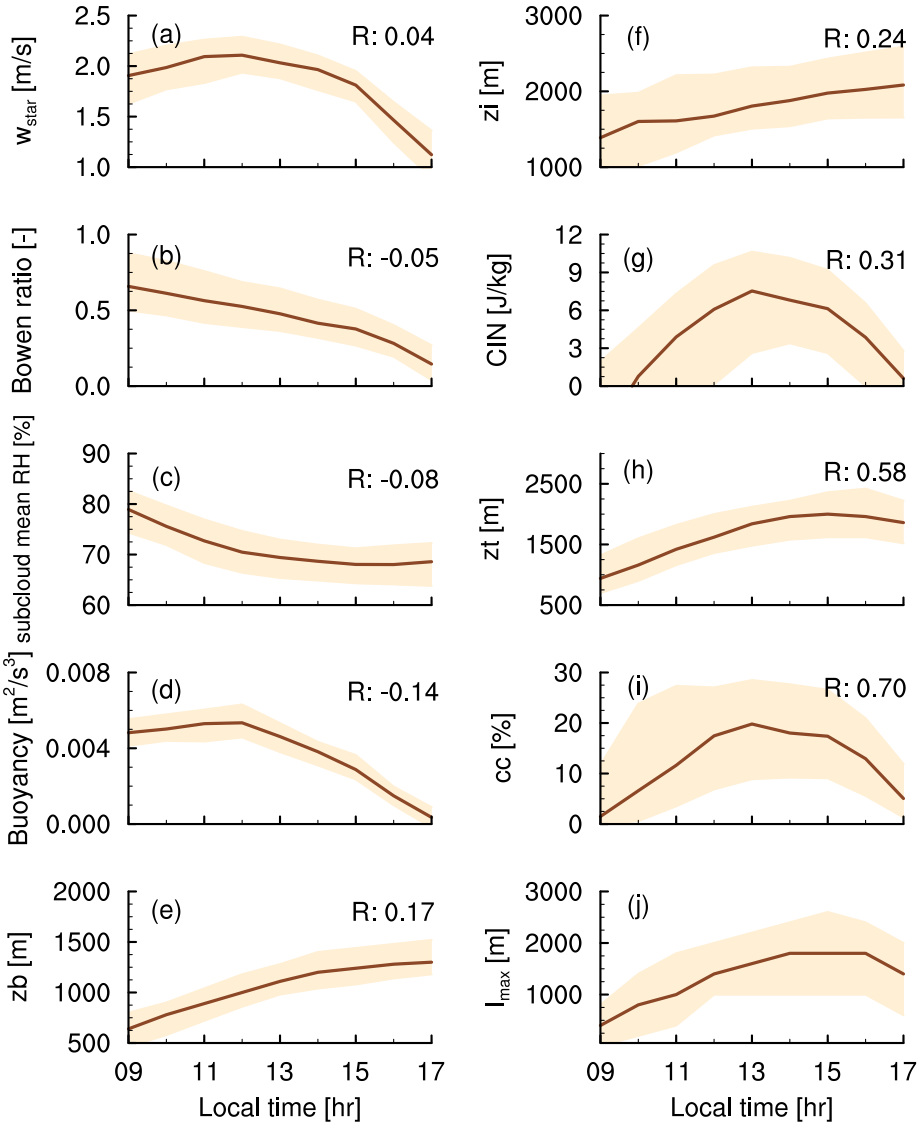

Figure 3.11: Temporal evolution of a selection of variables. The correlation coefficient (upper right corner of the panels) shows the correlation between the depicted variable and the maximum cloud size (panel h).

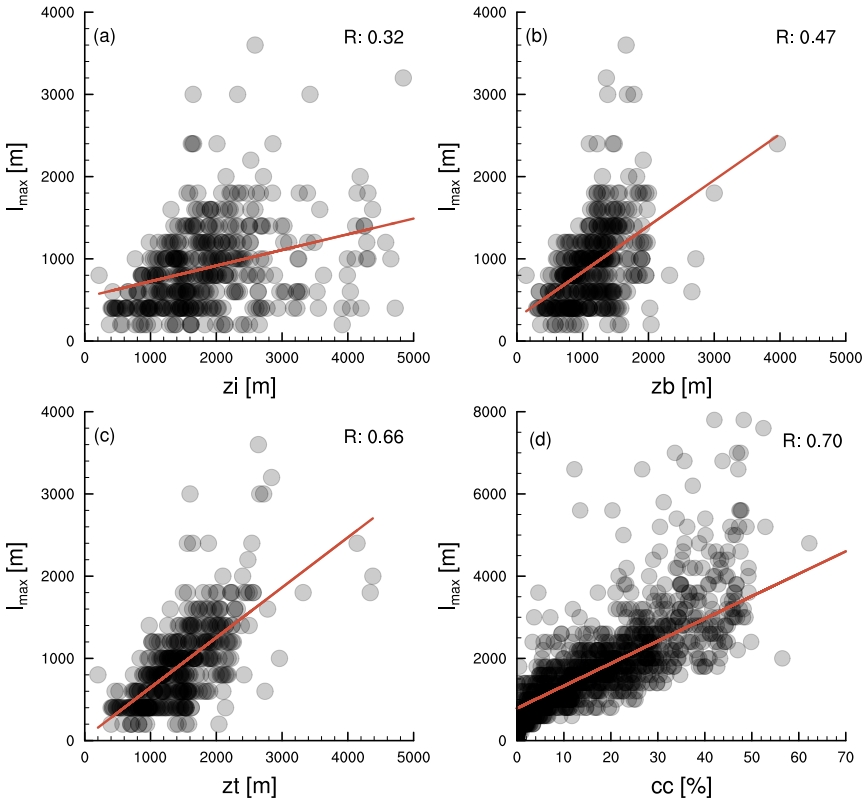

Figure 3.12: Scatter plots of $l_{\max}$ and cloud base, cloud top, boundary layer height and cloud cover. For cloud base, cloud top and boundary layer height only the days with cloud covers up to 10 % are taken into account. The correlation coefficients (R) can also be found in Table 3.4.

also compares well with the observed cloud chord length as shown in Zhang et al. (2017), even though the observations are taken at a different location.

3.5.4 The influence of cloud cover

To further disentangle the possibly simultaneous impact of multiple variables on $l_{\max}$, the results are stratified on the strongest correlating variable, which is cloud cover. Stratifying the data on the dominant dependence can act to highlight any secondary controls on the $l_{\max}$. To this end, the data is divided into bins of 10 % cloud cover. With this bin size we maintain sufficient statistical quality (more than 400 points for a cloud cover up to 10 % and around 200 for the other cloud cover bins), while narrowing the bin width enough to notice possible patterns. Resulting correlation coefficients are shown in Table 3.4, with scatter plots for the bold values in Figure 3.12a,b,c. The correlation between $l_{\max}$ and zi/zt/zb/CIN is stronger for a cloud percentage of 0-10 % than for all data together. For the higher cloud covers the correlation coefficients drop for zt, whereas they stay relatively constant for zi and zb. The correlation coefficients for CIN decrease most and show no strong correlation between CIN and $l_{\max}$ for the larger cloud covers. For low cloud covers the correlation coefficient is highest for zt. Figure 3.12c illustrates this high correlation coefficient, there are few outliers and no offset at the y-axis. Remarkably, the correlation coefficient for zi is lower than for zt and zb. However, it should be noted that even the 146 days of data might be insufficient for reliable correlation coefficients. The computed correlation coefficient can still be influenced by some outliers, as can be seen for example

Table 3.3: The correlation coefficient for the variables that show the strongest correlation with $l_{\max}$, for different cloud and cloud size definitions.

			Buoyant	
	Projected	Height-averaged	Projected	Height-averaged
zi	0.24	0.18	0.35	0.29
zt	0.58	0.5	0.53	0.42
zb	0.17	0.19	0.07	0.09
cc	**0.7**	0.7	0.72	0.65

Table 3.4: The correlation coefficient for projected maximum cloud size and boundary layer length scales, organized by different cloud covers.

cc [%]	zi	zt	zb	CIN
0-10	**0.32**	**0.66**	**0.47**	0.38
10-20	0.29	0.59	0.5	0.26
20-30	0.32	0.48	0.5	0.03
30-40	0.42	0.35	0.45	0.22

in Figure 3.12b. Nevertheless, the results suggest that after cloud cover, the depth of the cloud layer which is represented by cloud base and cloud top, plays a secondary but still important role in establishing the maximum cloud size in the population.

The four variables showing the best correlation with $l_{\max}$ are now investigated more closely. The variables considered are boundary layer height, cloud top, cloud base, and cloud cover. To assess the robustness of these high correlations, the calculation is repeated using slightly different definitions for a cloud and its size as described in section 3.33.3.1. In Table 3.3 the correlation coefficients are given between all five variables and $l_{\max}$, and for both definitions of a cloud and its size. For all four definitions a strong correlation exists between $l_{\max}$ and cloud cover. When taking only the cloud core into account, the correlation for all variables decreases slightly or stays the same. The difference between projected and height averaged cloud sizes is also small. There is a slightly better correlation for zb, but for the other variables there is a small drop in the correlation coefficient. The high correlation coefficient for the projected $l_{\max}$ and cc is visualized in Fig. 3.12d.

The strong correlation between $l_{\max}$ and cloud cover is robust for all tested cloud and cloud size definitions, as seen in the last row of Table 3.3. This could mean that fields with larger cloud cover feature larger clouds, instead of a larger number of smaller clouds. To test this hypothesis we look at the cloud cover decomposition $\alpha(l)$:

$$\alpha(l) = \frac{l^2 \; N(l)}{L_x L_y}, \tag{3.4}$$

here L_x and L_y are the horizontal dimensions of the domain (Neggers et al., 2003). The cloud cover decomposition gives the contribution to the total cloud cover as a function of cloud size. To answer the question of the relative contribution of larger clouds to high cloud covers, we again divide the data based on cloud cover in bins of 10 %. For every cloud cover bin the cloud cover decomposition $\alpha(l)$ is shown in Figure 3.13.

All cloud covers show a decreasing contribution to total cloud cover with increasing cloud

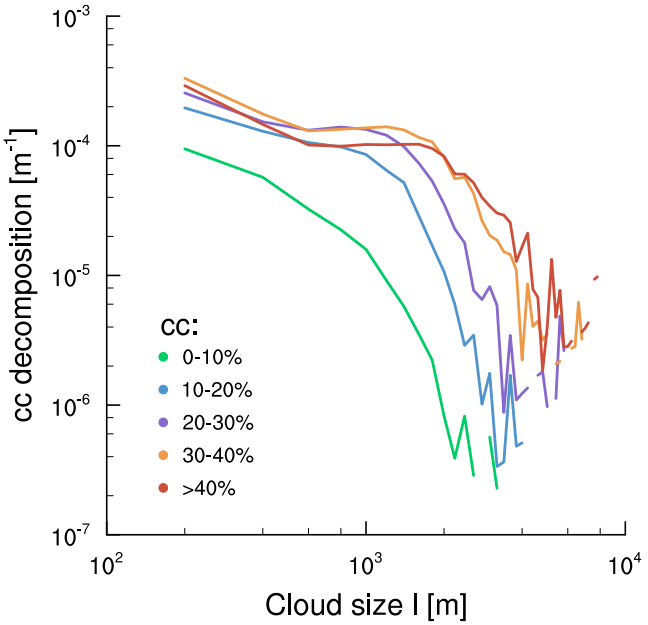

Figure 3.13: Cloud size against the cloud cover decomposition (Eq. 3.4). The data is divided into bins with similar cloud covers.

size. This decrease is stronger for smaller cloud covers whereas the slope is less steep for larger cloud covers. The small cloud sizes contribute less to the total cloud cover for fields with cloud covers up to 10 % as compared to larger cloud covers. The differences for larger cloud covers are minimal in this range. This is different for larger cloud sizes, here an increase in contribution is visible when going from small to larger cloud covers. These findings confirm that an increase of cloud cover goes together with an increasing contribution of large clouds instead of a larger number of small clouds.

3.5.5 Impact of microphysics

The way condensation and rain formation are described by a microphysics scheme in LES could in principle affect cloud sizes and their distribution (Zhang et al., 2017)). This is investigated by using an extra set of simulations of the days in 2013 and 2014 with the microphysics scheme of Sommeria (1976), from here on referred to as the Sommeria-Deardorff scheme. This is a scheme often operated by early LES models (e.g. Siebesma and Cuijpers (1995)). The simulations in this paper used for analysis are done with the Seifert-Beheng scheme (Seifert and Beheng, 2004). Independent of the microphyiscal scheme the all-or-nothing approach for cloud cover is used. This defines a grid box as cloudy once it is completely saturated. The main difference between the schemes is that in the Seifert-Beheng scheme rain is included and that it features number as well as mass densities. The formation and evaporation of rain has an impact on the heat and moisture budgets of the boundary layer. This could in turn influence cloud formation processes, and thereby cloud sizes. This short sensitivity test focuses on the first order impact of the microphysics scheme on cloud sizes. Individual components of the schemes are not addressed since this is beyond the scope of this study.

For both microphysics representations the diurnal evolution of the shape of the CSD as expressed by the hourly averaged fitting constants b and c is shown in Figure 3.14. The diurnal evolution for both constants is similar for the two microphysic schemes, in the morning b becomes smaller and c bigger. There is a small difference between the two

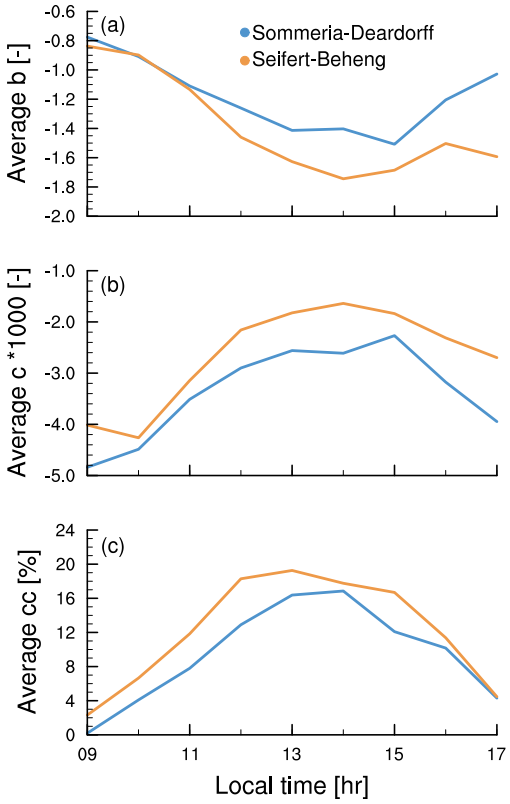

Figure 3.14: The diurnal evolution of the fitting constants b (panel a) and c (panel b) according to the power law-exponential function for two different microphysic schemes. The accompanying cloud cover is shown in panel c.

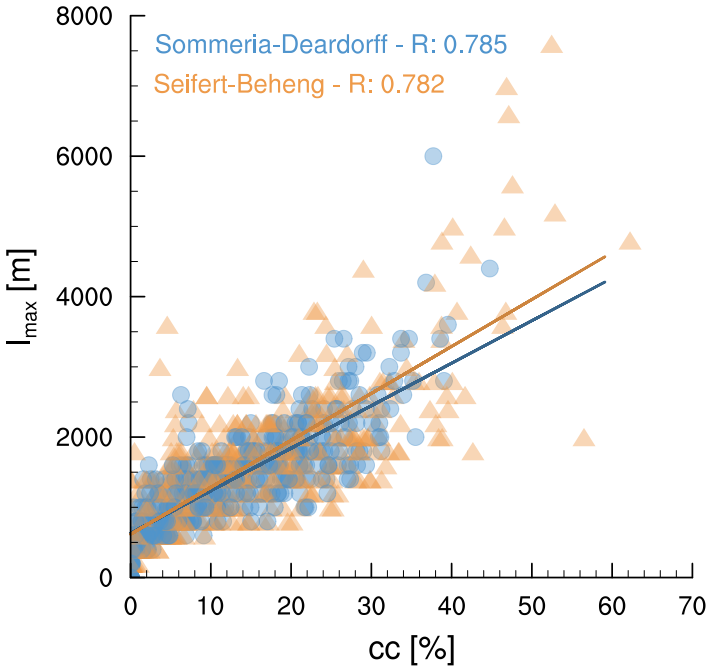

Figure 3.15: Scatter plots of $l_{\max}$ and cloud cover for two different microphysic schemes. The blue circles represent the Sommeria-Deardorff scheme, the orange triangles the Seifert-Beheng scheme. The correlation coefficient (R) for both sets of data is depicted in the upper left corner.

schemes, the b reaches lower and the c reaches higher values for the Seifert-Beheng scheme as compared to the Sommeria-Deardorff scheme. This combination of fitting constants is associated with a higher cloud cover, which can indeed be observed for the diurnal evolution of cloud cover shown in Figure 3.14c. Besides these small differences, the signal in the diurnal evolution for the shape of CSD is the same for both microphysic schemes.

Secondly, we test the strong correlation we found between $l_{\max}$ and cloud cover for both microphysic parametrizations. Again, using the data from 2013 and 2014, the correlation coefficients are computed between the hourly averaged $l_{\max}$ and cc. This results for the Sommeria-Deardorff scheme in an R of 0.785, while for Seifert-Beheng the R is 0.782. Also the scatter plots with regression lines show only small differences (Fig. 3.15). Both regression lines show a similar slope, the Seifert-Beheng line is a bit steeper. Probably due to the more advanced microphysics the $l_{\max}$ is slightly bigger for these simulations. To summarize, the correlation found between $l_{\max}$ and cc is according to these results not dependent on whether the Sommeria-Deardorff or the Seifert-Beheng microphysic parametrization scheme is used.

3.6 Discussion

The main goal of this study is to quantify the diurnal evolution of the CSD for continental shallow cumulus with a focus on its shape and range.

Concerning the range of the CSD, we find high correlations between $l_{\max}$ and cloud cover, cloud base and cloud top, and less so with the boundary layer height. The latter is surprising, because it is classically used as a scale that controls many aspects of the boundary layer. It is often used to normalize height and is seen as a measure for the vertical scale of

the largest thermals. It would thereby control the maximum horizontal cloud size as well. However, here evidence is found that other controlling factors exist which co-determine the size of the cloud. The fact that correlation coefficients for cloud cover, cloud base, and cloud top are higher than for boundary layer depth suggests that there might be another factor that eventually effectively controls the cloud size. This also means that it might be necessary to reconsider the strong assumption that is used in some cumulus parametrizations.

A possible physical explanation for the high correlation between cloud cover and $l_{\max}$ could be the dryness of the cloud environment. The small cloud covers (up to 10 %) can be seen as forced cumuli, they reach the inversion and just condense, but their environment is very dry, preventing them from growing further, both in the vertical and the horizontal. When the environmental air that surrounds the rising cumuli at the top of the boundary layer is sufficiently humid or close to saturation, the stronger cumuli manage to condense a much larger part of their area, not just their core. In addition, clouds dilute less quickly in more humid air, prolonging their lifetime and promoting the formation of outflow or 'anvil' clouds. Further research is necessary to investigate this relation between the moisture content of the environmental air and the cloud cover/$l_{\max}$. Some first steps were taken on this, but addressing the humidity in the direct surroundings of the cloud is not so straightforward since this really depends on the height and structure of the cloud. It is therefore considered outside the scope of the present study and topic for further research.

The strong correlation between $l_{\max}$ and cloud cover reveals something important about the response of a cumulus cloud population to a diurnally varying external forcing. We find that an increase in cloud cover is not generated by a larger number of small clouds, but is instead carried by larger clouds. This behaviour is relevant for radiation schemes in GCMs. The effective vertical transfer of radiation depends on the sizes and horizontal distribution of clouds (Barker and Räisänen, 2005). Organising the data by cloud cover showed a secondary role for cloud base height and cloud top height in controlling $l_{\max}$. Although strong correlations do not reveal anything about causality, this does indicate that besides cloud cover, the heights of cloud base and cloud top play a secondary but still important role in establishing the maximum cloud size in the population.

The study by Sakradzija and Hohenegger (2017) suggested a strong link between the surface Bowen ratio and the mass flux distribution. The low correlation between $l_{\max}$ and the Bowen ratio reported in this study does not necessarily contradict that finding. It rather suggests that the vertical velocity component of the mass flux could carry the strongest dependence on the Bowen ratio (Lamer and Kollias, 2015), and not the area of the largest thermal. To prove this hypothesis requires further research.

3.7 Conclusions

To improve our understanding of what controls the shape and range of cumulus cloud size distributions, we performed an ensemble of LES simulations of 146 cases of diurnal cycles of summertime shallow convection over land at a mid-latitude meteorological site in north-western Germany. To this purpose, the relation between defining aspects of the CSD, including the slope and the maximum size, with various candidate variables expressing boundary layer state was investigated.

The power law-exponential function, which was used to fit the cloud size distributions, shows the typical slope which has been reported in previous studies. This slope covers a larger range of cloud sizes as cloud cover increases. The shape of the CSD changes significantly during the day, with a shift from exponential at cloud onset to power law

at midday and back to exponential in the evening. What this means in terms of the spatial structure of the cumulus cloud field, including effects of merging of clouds and neighbour spacing, is not explained by the results obtained in this study. Future research will therefore focus on the interaction between clouds in order to better understand the inter-cloud dynamics. A possible way to do this is to determine the cloud spacing in relation to cloud size and its relation with cloud cover.

A dominating correlation between cloud cover and maximum cloud size is found, although a weaker secondary relation with cloud base height and cloud top height is also identified. The correlation between cloud cover and maximum size implies that it is not a larger number of small clouds, but rather larger clouds that are responsible for an increase in cloud cover. The possible role of environmental humidity in the strong link between cloud cover and maximum size needs to be investigated further.

With a subset of the data sensitivity tests are performed on the output resolution and the microphysics scheme used in the LES. Both sub-sampling of the original resolution and the use of a less involved microphysics parametrization have no influence on the conclusions of this study.

Results of this study showed that the shape of the CSD is related to the cloud cover. This could explain the different CSDs found in previous literature. Insights like this are beneficial for the development of scale-aware parameterization schemes based on CSDs, they need to be informed about the shape of the CSD for closure and calibration (Wagner and Graf, 2010; Neggers, 2015; Brast et al., 2018).

Acknowledgments

This research is supported by the SFB-TR32 “Patterns in soil-vegetation-atmosphere systems: monitoring, modelling and data assimilation” funded by the German Science Foundation (DFG). The authors are thankful for given the opportunity to run simulations on CHEOPS, the cluster of the Regionales Rechenzentrum of the University of Cologne as well as on the supercomputer JURECA at Jülich Supercomputing Centre (JSC). Observations used in this study originate from German Research Foundation (DFG) funded Core Facility (JOYCE-CF) under DFG research grant LO 901/7-1. The observational data used in this study is available at the Institute for Geophysics and Meteorology of the University of Cologne. Last, the authors want to thank Steve Klein and two anonymous reviewers for their useful comments and help in improving this paper.

3.8 Appendix: comparison of the original and coarse resolution DALES output

For computational reasons only every fourth point of the original LES output is used in this study, resulting in an effective horizontal resolution of 200 m. However, the simulations are done with a resolution of 50 m. In this section we will evaluate the consequences this sub-sampling of the original data might have.

We only take into account the 52 simulated days of the years 2013 and 2014. As a first step to see the influence of resolution on the CSD, the normalized distributions are computed for a single moment, at 13:00 LT on 04-07-2013 (Fig. 3.16). Through the data (dots) the power law-exponential is fitted (lines). Green denotes the sub-sampled data (as used in this study), and shows a power law for the smaller sizes with an exponential for the larger sizes. The original resolution (red) displays a power law across all sizes. It seems that this distribution is heavily influenced by the smaller cloud sizes. For a fairer comparison the fit is therefore also determined by using the same cloud size as starting point (orange line). This adjusted fit is based on the data for the original resolution, but ignoring cloud sizes smaller than the smallest bin of the sub-sampled data. Now again the exponential part is seen at the bigger cloud sizes. This is supported by the values of the fitting constants, as shown in Table 3.5. Since for the original resolution clouds smaller than 200 m are included as well, the total number of clouds is larger. This leads to a lower normalized CSD, as is also visible in Figure A1. For an easier comparison of the slope and shape of the different fits, the fit through the sub-sampled data is plotted again in dotted grey. This is done in such a way that the two different densities match at the smallest bin of the coarse resolution.

To study the temporal evolution of the shape of the fit for both resolutions, all days of 2013 and 2014 are used. The averaged b and c constants of the fit are plotted against time (Fig. 3.17a,b), with colours corresponding to the ones in Fig. A1. It can be seen that no matter the resolution, the temporal evolution of the fitting constants is similar. The b

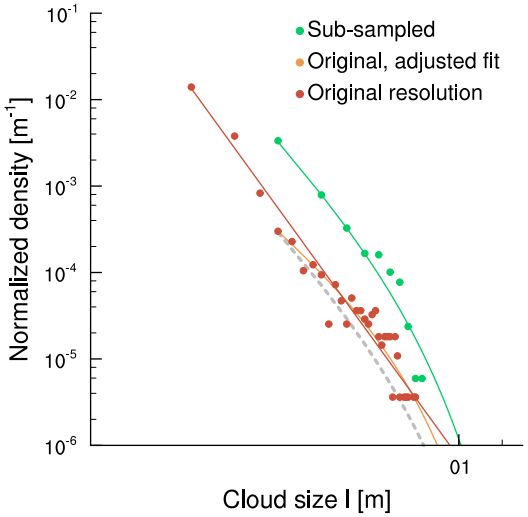

Figure 3.16: Cloud size densities for the original and coarse resolution (dots), including fits according to the power law-exponential function (lines). Densities are for 13:00 LT at 4 April 2013. The grey dotted line is the reflection of the fit through the sub-sampled data, only plotted in such a way that the densities match at the smallest bin of the coarse resolution.

Table 3.5: Fitting constants for CSD for 4 April 2013 for different output resolutions. The CSD are shown in Figure A1.

	a	b	c (x 1000)
Original resolution	108.97	-2.3	0.0
Original, adjusted fit	0.46	-1.35	-1.0
Coarse resolution	52.78	-1.8	-0.8

decreases during the day and the c increases, this process reverses at the end of the day. The largest differences between the resolutions can be seen in the early morning hours. For b, the adjusted fit for the original resolution is closest to the coarse resolution during the afternoon. For c, this is the case for the unadjusted original resolution. The maximum cloud size is an important variable in this study and is therefore also tested for influence of sub-sampling. During the whole day, the averaged l_{max} as simulated by DALES (Fig. 3.17c), is very similar to the original resolution and the sub-sampled one, with an increase in the morning, a peak during mid-afternoon and a small decrease afterwards.

The magnitude and evolution of the b and c fitting constants of the power law-exponential fit is similar for both resolutions, also the l_{max} is not influenced by sub-sampling. Therefore, the use of the coarse resolution for the analysis in this paper is justified.

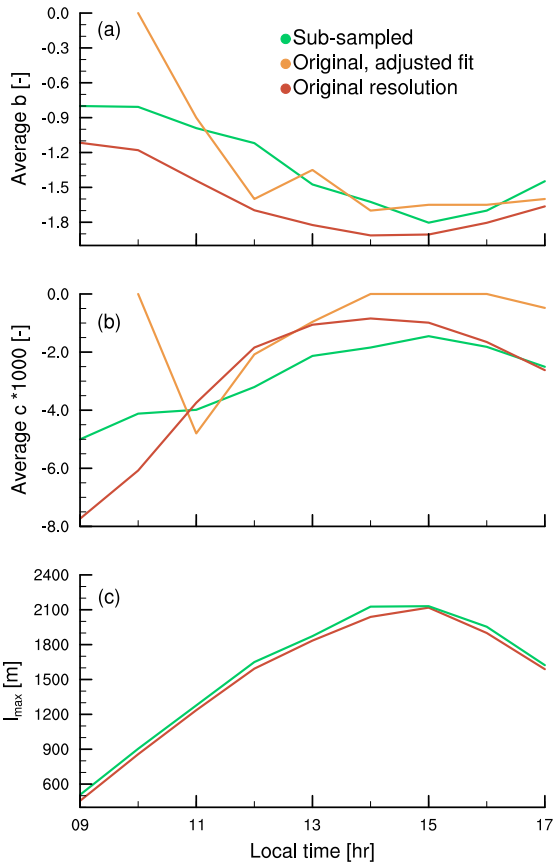

Figure 3.17: Evolution of the fitting constants b (a) and c (b) following the power law-exponential function, and the $l_{\max}$ from DALES (c). They are averaged over the CSDs at each timepoint, for the original and coarse resolution.

Chapter 4

On the size dependence of cumulus cloud spacing

This chapter is part of: Thirza W. van Laar and Roel A.J. Neggers (2019), On the size dependence of cumulus cloud spacing. *Journal of Geophysical Research: Atmospheres*, submitted

Abstract

In this study the spatial organization of Trade wind shallow cumulus populations as diagnosed from large domain LES is investigated, with a special focus on the implications for cumulus parametrization in the grey zone of convective parametrization. For this, various expressions of the nearest neighbor spacing for shallow cumulus are studied, exploring in particular how cloud spacing depends on cloud size. Four different definitions are used, these include the spacing between clouds of any size, the spacing between clouds of similar size and using either cloud edge spacing or cloud center spacing. We report a different relation between cloud size and spacing for different definitions; the relation is exponential for the spacing between clouds of similar size and logarithmic for the spacing between clouds of any size. A simple conceptual model is formulated to explain this behaviour, suggesting that randomly distributed small clouds can cause the logarithmic dependence.

Plain language summary

Fair-weather, or shallow cumulus clouds, persistently cover a large part of the earth, especially in the marine subtropics. These small clouds play an important role in the energy balance of the atmosphere because they reflect radiation, an effect that needs to be accurately represented in weather and climate models. Two recent developments have prioritized the scientific research into the spatial distribution of cumulus clouds. These include i) the insight that the spatial organization of cloud fields affects climate feedbacks, and ii) the ever increasing resolution of weather and climate models. To gain further insight, in this study we analyse large fields of shallow cumulus clouds to investigate how their spatial distribution depends on their size. We find that cloud spacing increases with cloud size. The functionality depends on the definition of cloud spacing and is either logarithmic or exponential. A simple conceptual model is formulated to explain the logarithmic dependence, suggesting that random clustering of smaller clouds around larger ones is causing this behavior.

4.1 Introduction

One of the challenges of present-day climate and weather prediction modeling is the grey zone problem (Wyngaard, 2004), which means that at present-day resolutions, previously parameterized processes are becoming partially resolved. To address this problem, there is an ongoing development of scale-aware parametrization schemes for boundary layer-scale processes, such as the formation of shallow cumulus clouds. A potential way forward is the formulation of spectral or PDF convection schemes that are based on a discretized cloud size distribution (CSD) (Wagner and Graf, 2010; Park, 2014; Neggers, 2015). By incorporating a CSD, size information is directly built into the scheme, which allows filtering on scale (Brast et al., 2018) and representing impacts of spatial organization (Neggers et al., 2019). The spatial organization of convection affects the mean state of the atmosphere and thereby climate feedback mechanisms and is therefore important to take into account (Bretherton et al., 2005; Wing et al., 2018).

Since the CSD fulfils an important role in such schemes, it is necessary to understand what factors control its behavior. For this, information is needed on its shape, range and variability. These aspects are influenced in two ways. First, there is the general influence of the large-scale meteorological conditions, which affect all clouds similarly. Second, clouds also interact with each other, which is a form of population dynamics (Lotka, 1920; Volterra, 1926). Cloud spacing is an expression of these interactions. Therefore, a good starting point when studying cloud interaction is Nearest Neighbor Spacing (NNS).

NNS has played a key role in previous research of the behavior of cumulus cloud populations. Based on the cumulative distribution of the NNS, a method has been developed to assess organization in a cloud field (Tompkins and Semie, 2017; Weger et al., 1993; Sengupta et al., 1990). NNS has a direct application in the development of scale-aware stochastic parameterization schemes (Neggers et al., 2019). NNS becomes relevant in the grey zone of convection (Wyngaard, 2004), when the model grid spacing becomes comparable to the cloud spacing. This cloud spacing plays a crucial role in explaining the powerlaw scaling in the internal variability of the CSD. For further development of these schemes, a better understanding of NNS is needed.

Of particular interest is how NNS depends on cloud size. Early insights were provided by Joseph and Cahalan (1990), who found a positive linear relation between cloud size and NNS. That result indicates that the size of a cloud impacts its surroundings in such a way that larger clouds have a bigger spacing. They analysed satellite snapshots of cumulus clouds taken all over the globe. For all these snapshots, the meteorological and surface conditions were different, which might be the cause of the identified different relations between NNS and cloud size per snapshot. Two further studies on the spatial structure of cumulus clouds that use satellite observations include Weger et al. (1992) and Nair et al. (1998). Both studies use the cumulative distribution function of NNS to assess spatial organisation in a cloud field.

In this study we employ large-domain simulations at high cloud resolving resolutions to gain more insight into cloud spacing in Atlantic Trade wind cumulus populations. LES can well be used to this purpose, as it provides complete 3D fields at high temporal resolutions (Neggers et al., 2003; Heus and Seifert, 2013; Rieck et al., 2014; Senf et al., 2018). The large domain ensures a large enough sample size to achieve statistical significance. The location over the ocean ensures fairly homogeneous conditions concerning the state of the atmosphere and surface characteristics. In the context of the project High Definition Clouds and Precipitation for Advancing Climate Prediction (HD(CP)^2) simulations were perforemd on a large domain with high resolution, which is necessary to study the spatial distribution of clouds. We examine the behavior of various possible definitions of the NNS,

focusing on the question how it depends on cloud size. Finally, a simple conceptual model is proposed to explain the functional relation between cloud spacing and cloud size.

The data and methods used in this study are described in the second section. How cloud spacing depends on cloud size is shown in section three, and a simple conceptual model for explaining the logarithmic size-dependence is formulated in section four. We end with some concluding remarks in section five.

4.2 Data and methods

4.2.1 ICON simulations

The data used for the analysis were generated by the HD(CP)^2 project with the Icosahedral Nonhydrostatic (ICON) model (Zängl et al., 2014; Heinze et al., 2017). The domain is located over the tropical Atlantic, where shallow cumulus clouds occur persistently year-round. The simulated day is 20 December 2013. The set-up is similar to that of Klocke et al. (2017), which consists of a set of four one-way nested domains at increasing resolution. Their 1.2 km resolution domain is our outer domain. At its boundaries the outer domain is forced by ECMWF data and 3 nests are added, with resolutions of about 600, 300 and 150 m. The most inner domain is used for this analysis, located directly to the east of Barbados and spanning approximately 150x400 km. The domain covers 21 km in the vertical, divided in 150 levels. With a horizontal resolution of 150 m the resolution of the inner domain is high enough to switch off all parametrization schemes except the ones for turbulence, cloud microphysics, and radiation.

The simulation starts at 12 UTC and ends 12 hours later. From these 12 hours, the first six are not considered because of spin-up effects. Every 15 minutes, a 3D field of liquid water is available as output, which serves as input for the clustering algorithm (as described in the next subsection). A snapshot of vertically projected liquid water (Fig. 4.1) shows that the domain contains numerous resolved clouds, up to 4500 per snapshot. This dataset can therefore be considered as a useful source to study spacing between clouds. Since the surface conditions are relatively homogeneous, accounting for influences from a heterogeneous surface on boundary layer development and cloud formation is not necessary and any organization will be due to large-scale effects or cloud-cloud interaction.

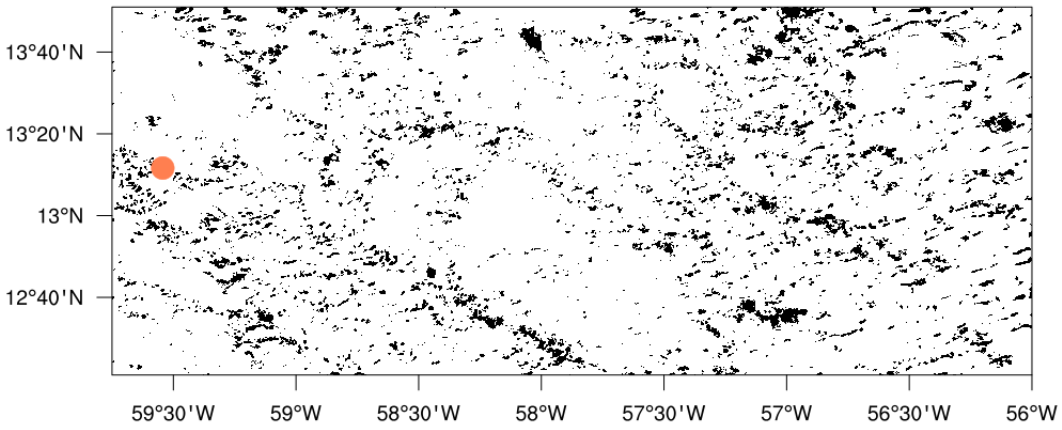

Figure 4.1: Snapshot of ICON output of projected cloud liquid water in black, for 20:05 UTC. The orange dot shows the location of the island of Barbados.

4.2.2 Clustering algorithm

A clustering algorithm is used to compute the cloud sizes and locations from the model output (van Laar et al., 2019). The liquid water field is projected on the surface and a grid cell is considered cloudy if the total liquid water content of the column is bigger than $1 * 10^{-8}$ kg/kg. If two cloudy cells share a cell edge, they are considered part of the same cloud. Cloud size is defined as the radius of a circle that has the same area of the cloudy grid cells belonging to the cloud (Rieck et al., 2014). These cloud sizes are distributed over bins with a width similar to the grid resolution (about 150 m). The center of mass of the cloud is taken as the center of the circle, the coordinates of this point are used for determining the spacing between the clouds.

4.2.3 Nearest Neighbor Spacing

In this study the method proposed by Joseph and Cahalan (1990) and Tompkins and Semie (2017) is adopted to calculate the NNS. It is defined as the distance from one cloud to its closest neighbor. This means in practice that, for every cloud, the minimum distance is selected from the distances to all other clouds. Let $\mathcal{K}$ represent the total set of clouds, with n the total number of clouds: $\mathcal{K} = \{1, 2, ..., n\}$. NNS between cloud k and its neighbors n is defined as:

$$NNS(k) = min\{d(n,k) \mid n \in \mathcal{K} \setminus \{k\}\}, \tag{4.1}$$

with d the great circle distance (euclidian distance corrected for the curvature of the earth). Alternative definitions of cloud spacing can also be thought of. For example, a distinction is made between *NNS* and the *equal-size* NNS (NNS_σ), where for the latter only distances between clouds of a similar size (l) σ are considered. This makes our set of clouds dependent on l: $\mathcal{K}_\sigma = \{k \in \mathcal{K} \mid l(k) = \sigma\}$. NNS_σ is then defined as:

$$NNS_\sigma(k) = min\{d(n,k) \mid n \in \mathcal{K}_\sigma \setminus \{k\}\}. \tag{4.2}$$

To calculate the spacing between the clouds, two different approaches were followed. First, the *cloud center spacing* is used, the distance from cloud center to cloud center. Second, the *cloud edge spacing* is the distance from cloud edge to cloud edge, computed by assuming that all clouds are perfect circles (Rieck et al., 2014; Dawe and Austin, 2013). In essence, it is the cloud center spacing minus the size (radius) of the two neighbouring clouds: $d_{CE} = d(n,k) - r_n - r_k$. Then the NNS_{CE} for using cloud edge spacing is defined as:

$$NNS_{CE}(k) = min\{d_{CE}(n,k) \mid n \in \mathcal{K} \setminus \{k\}\}, \tag{4.3}$$

and the equal size NNS using cloud edge spacing ($NNS_{\sigma,CE}$) as:

$$NNS_{\sigma,CE}(k) = min\{d_{CE}(n,k) \mid n \in \mathcal{K}_\sigma \setminus \{k\}\}. \tag{4.4}$$

4.3 Results

The four panels of Figure 4.2 show the size-dependence of all four NNS definitions as defined in the previous section. The results are averaged over all snapshots. The R^2 reflects the quality of the fit. Fig. 4.2a shows the size-dependence of NNS, which in the range < 600 m

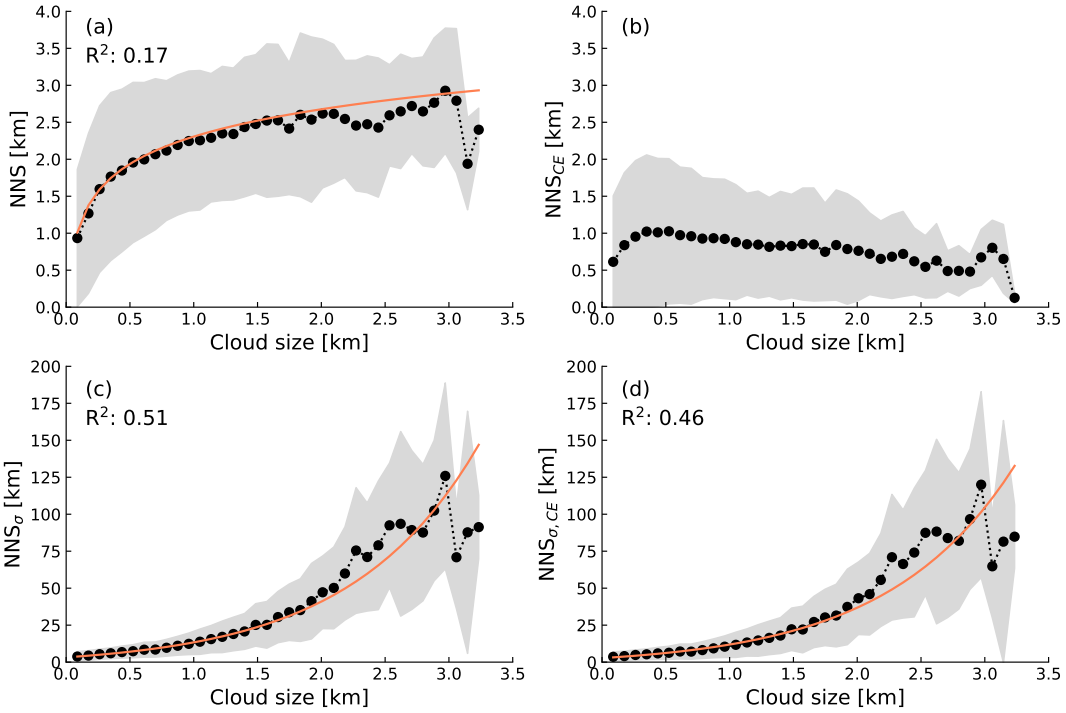

Figure 4.2: Discretized cloud size against averaged a) NNS, b) NNS_{CE}, c) NNS_{σ}, and d) $\mathrm{NNS}_{\sigma,CE}$. The NNS is averaged over all analysed fields, the grey shading shows the mean $\pm$ the standard deviation. The orange lines show the best fits through the data with their R^2 value in the upper left corner.

roughly matches the linear fit reported by Joseph and Cahalan (1990). However, at larger sizes the dependence is best captured by a logarithmic relation ($y = 2.31 + 1.23\ \log_{10}(x)$), with an R^2 value of 0.17. For the larger cloud sizes, the mean falls slightly below the fit. The limited amount of data might be a reason for this; the larger the cloud size, the fewer clouds of that size are present in the field. A statistical analysis revealed that NNS is underestimated for fewer than 10 clouds per cloud size (not shown).

NNS_{σ} is shown in Fig. 4.2c, reflecting NNS when only the spacing of clouds of a similar size are taken into account. Again we find a positive relation between cloud spacing and cloud size. This time, however, the relation is best captured by an exponential ($y = -2.66 + 5.90\ exp(x)$) with an R^2 of 0.51. Other differences with NNS are i) the larger spacing and ii) an increasing spread around the mean. A possible explanation is that clouds of the same size do not cluster, but are homogeneously distributed over the domain. This spread then leads to larger spacing, especially for bigger clouds, which are not so numerous.

Fig. 4.2b and d show NNS_{CE} and $\mathrm{NNS}_{\sigma,CE}$, respectively. When interpreting these results it is important to consider that the spacing for bigger clouds could be larger simply because their centres are spaced further apart, due to their size. Therefore these two NNS definitions now reflect the Cloud Edge (CE) spacing. This slightly different definition of cloud spacing leads only to very small differences for NNS_{σ}, only the spacing is smaller (Panel d). The exponential relation is present again ($y = -2.60 + 5.33\,exp(x)$), albeit with a slightly lower R^2 value of 0.46. However, for NNS_{CE} (Panel b), the clear logarithmic relation with cloud size like for NNS is no longer present. After a first increase of NNS_{CE} with cloud size, for clouds larger than about 400 meter a slight decrease of spacing with size is visible.

The difference between cloud-edge and cloud-center spacing can be understood by consid-

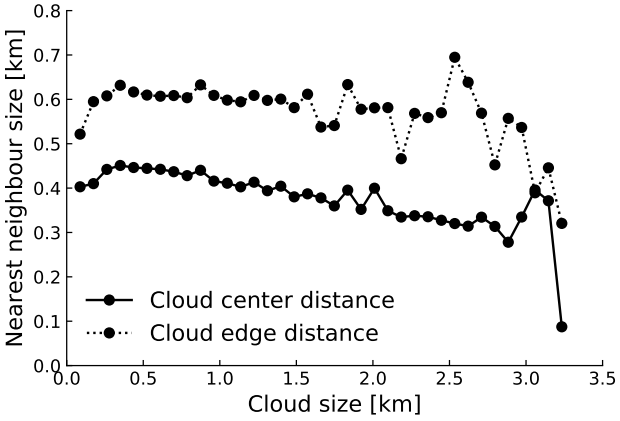

Figure 4.3: Discretized cloud size against the averaged nearest neighbour size, using both the cloud center spacing (solid) and the cloud edge spacing (dotted).

ering the size of the nearest neighbours (Fig. 4.3). The size of the nearest neighbour reflects the spatial distribution of clouds of different sizes relative to each other. Nearest cloud size is considered as a function of the cloud center spacing and the cloud edge spacing. For both methods, after a slight increase for the small cloud sizes, the size of the neighboring clouds slowly decreases with cloud size. This behavior suggests that larger clouds are increasingly surrounded by smaller clouds. What also stands out is a pronounced difference between the sizes of the nearest neighbors for the two definitions of spacing. Although both definitions show a similar functional dependence, the averaged neighbor size is larger when using cloud edge spacing. This indicates that the chance that clouds have a bigger neighbor is higher when using cloud edge spacing. A simple thought experience can explain this, featuring one small and one big cloud. Using cloud center spacing would make the spacing between these two large, because of the radius of the big cloud. This large spacing makes it less likely that they are each other's nearest neighbor. However, the radius of the big cloud does not play a role anymore when using cloud edge spacing, hence their spacing is smaller and the chance that they are each other's nearest neighbor is larger. Therefore, using cloud edge spacing means that larger clouds can also have a small spacing to their nearest neighbor, which can explain the decrease of NNS_{CE} and the smaller spacing in Panel b from Figure 4.2 as opposed to Panel a.

4.4 A simple conceptual model for NNS

The logarithmic functional relation between cloud size and NNS could be related to how clouds of different sizes are spatially distributed with respect to each other. For example, an increasing spacing between bigger clouds makes it more likely that a smaller cloud is present in between. The NNS would then saturate for the larger cloud sizes, leading to a logarithmic functional relation. To gain insight a simple conceptual model is now formulated to explain this saturation effect.

The conceptual model is based on a population of clouds in 1D space. The population consists of two sizes of clouds, big ones and small ones. For simplicity, the big ones are assumed to have an equidistant spacing, while the small ones are randomly distributed. Every cloud is represented by its location, which is taken from a series of integers spanning 1 to 1000. An example of such a field can be seen in Figure 4.4a, where every clouds' location is represented by a line. The spacing between the larger clouds (orange lines) is assumed to increase linearly with size (Neggers et al., 2019), following the relation:

spacing $= 5 + 5 *$ size. For every size of the big clouds, the population is constructed 50 times, after which the results are averaged. In this configuration the NNS corresponds to the absolute value of the difference in location between a cloud and its nearest neighbour.

Figure 4.4b shows the resulting NNS as a function of the size of the larger clouds. The conceptual model indeed reproduces a saturating NNS with cloud size, in this respect reproducing the logarithmic behavior as diagnosed in the LES cloud fields (see Fig. 4.2a). While the linear increase in NNS for the small sizes is not captured well by the logarithmic fit (orange line, $y = 3.43 + 1.68 \log_{10}(x)$), the general function dependence for larger sizes matches well with the LES results. The reproduction of the logarithmic behavior indicates that the functional relation between NNS and cloud size is dominated by a random distribution of small clouds around larger ones.

4.5 Concluding remarks

We studied NNS as a function of cloud size for shallow cumulus clouds in marine subtropical Trade wind conditions. The results showed increasing NNS with cloud size. The functional relation depends on the definition of spacing, moving from NNS to NNS_σ changes the relation from logarithmic to exponential. A conceptual model is formulated that suggests that random distribution of small clouds around bigger ones can explain the logarithmic dependence. A nearest-size analysis suggests some clustering of small clouds around bigger ones takes place.

An increasing NNS with size thus means that the larger a cloud, the larger the chance it is close to a smaller cloud. This relation might be because of larger clouds falling apart but it could also be a form of organization since this influences NNS (Tompkins and Semie, 2017). Organization could also be an explanation for the exponential relation between cloud size and NNS_σ. The results show that the larger the cloud, the bigger the spacing to a cloud of a similar size. This suggests that the cloud is based on a bigger circulation cell. These circulation cells might point to some organizational structures, e.g. in the form of cold pools. The formation of cold pools depends on the evaporation of rain, and rain can only be formed if a cloud is big enough. Since organization in the form of cold pools affects

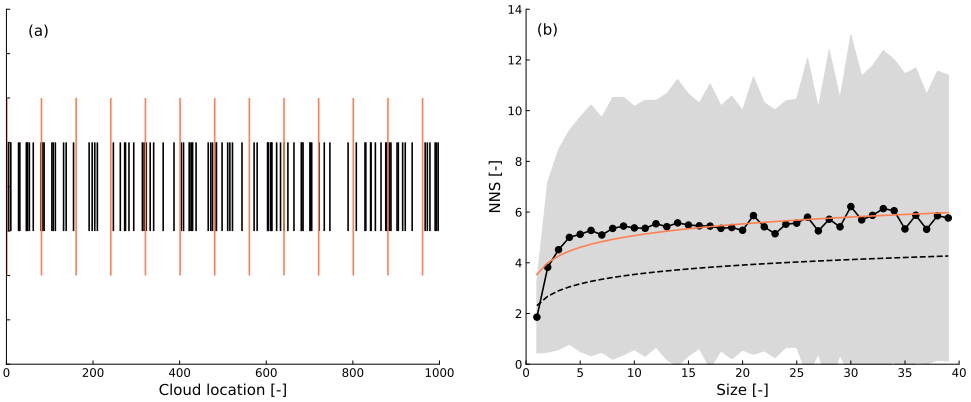

Figure 4.4: a) An example of a field of clouds. The black lines represent randomly distributed clouds with small spacing whereas the orange ones represent larger clouds with an equidistant spacing. b) Averaged NNS (black dotted line) ± the standard deviation (grey area, upper panel) as based on randomly generated clouds placed in a 1D field. The orange line shows a logarithmic fit through the data. The black dashed line is the fit from Figure 4.2a.

larger clouds most, cold pools could explain the exponential relation between cloud size and spacing. The exact relation between cloud field organization and cloud spacing is a topic of further research. The logarithmic behaviour is explained by the conceptual model we introduced. This model showed that the saturation of NNS with cloud size stems from randomly distributed small clouds around larger ones.

Our analysis of NNS proved to be useful in gaining insight into the relation between cloud size and spacing. Apart from the impact of organization, the results motivate further investigation of other aspects, including the influence of surface and large-scale conditions to be able to apply the findings in more general settings. An obvious next step is to apply our analysis to observed cloud populations, for example as detected by high-resolution satellite instrumentation. This is an important future research topic for assessing the validity of the results based on LES experiments as presented here.

Acknowledgments
This research is supported by the SFB-TR32 'Patterns in soil-vegetation-atmosphere systems: monitoring, modelling and data assimilation', funded by the German Science Foundation (DFG). The authors want to thank Matthias Brueck from the Max-Planck-Institut für Meteorologie who performed the ICON simulations in the scope of the HD(CP)^2 project. The data used in this study is available from the HD(CP)^2 project upon request.

Chapter 5

Quantification of organization in shallow cumulus cloud populations using large-domain LES

Abstract

Shallow cumulus clouds often occur in clusters and this organization impacts e.g. radiation and precipitation. To account for this effect in atmospheric models, an organization parameter needs to be included in parameterization schemes for convection. Since horizontal resolutions of atmospheric models keep increasing, a parameter is needed that can be used across a range of scales to be introduced into scale-aware parameterization schemes. Recently, several parameters have been developed to capture the spatial distribution of clouds in a single number. However, these parameters have mainly been evaluated using data for deep convective cases. In this study the parameters described in literature are therefore evaluated for a large-domain LES, located over the tropical Atlantic and featuring numerous shallow cumulus clouds. We compare a set of recently used parameters for expressing organization in convective cloud fields. Additionally, the unsupervised learning method hierarchical clustering is applied, since it could potentially yield an alternative parameter. All parameters are treated as similarity measures; they are evaluated against a baseline of randomness. The results show that three of the parameters are able to capture the transition from more organized (during spin-up) to less organized cloud fields. The hierarchical clustering method indicates that in most analysed cloud fields no significant clusters of clouds could be detected. By assessing the advantages and disadvantages of the organization parameters, it can be concluded that from the ones compared in this study the Organization Index (I_{org}) has most advantages.

5.1 Introduction

The grouping of shallow cumulus clouds into clusters is regularly observed, as is already documented by Plank (1969) and Randall and Huffman (1980). These early observations, both by eye and by satellite, motivated atmospheric scientists to study clustering, randomness, and regularity in cloud fields (Weger et al., 1992). With the development of numerical cumulus models (Hill, 1974) and the realisation that cumulus organization impacts the radiative budget and produces more intense rain (López, 1978), it became necessary to better understand and describe the organization of clouds. This was mainly done by using satellite data. Nair et al. (1998) provide an extensive overview of these early developments in studying the spatial distribution of shallow convective clouds.

Recently, high resolution simulations of clouds enabled by ever more powerful supercomputers are providing new insights into the spatial organization of cloud populations. It is understood that cloud-cloud interaction has to be taken into account to explain and understand convection (Haerter et al., 2017). Many numerical simulations show organization (Wing et al., 2017), even with the absence of surface heterogeneities, rotation or mean pressure gradients (Radiative Convective Equilibrium simulations), (Bretherton et al., 2005). The ongoing study of convective organization using numerical simulations shows the impact on atmospheric conditions. Organization has a drying effect and impacts radiation (Wing and Cronin, 2015). It might be affected by surface temperatures (Coppin and Bony, 2018), and influences precipitation (Cronin and Wing, 2017). The impact of convective organization on climate is a topic of discussion (Wing, 2019).

An evaluation of organization parameters for convection becomes especially important now that horizontal resolutions of atmospheric models keep increasing and the grey zone of convective parametrization is reached. The strict scale separation between small- and large-scale convection is not applicable any more with the increase of model resolution. A convection scheme that is able to operate in the grey zone therefore needs to be scale-aware. For the development of such scale-aware parameterization schemes, the effect of spatial organization needs to be taken into account (Neggers et al., 2019).

Several parameters have been developed to express the degree of organization. The organization parameters are able to categorize the spatial distribution into regular, random or clustered, or allow intercomparison of the degree of organization. The latter are difficult to compare across different datasets, due to their dependence on specific conditions like resolution and domain size. Most organization parameters are applicable to both observational as well as simulation data, allowing a direct comparison between the two. The organization parameters have mainly been studied and applied for deep convection. Patterns in stratocumulus have been studied as well, using natural cellular networks (Glassmeier and Feingold, 2017). Studies of organization have mainly been limited to deep convection. In a key study Bretherton and Blossey (2017) were the first to investigate this phenomenon in shallow cumulus cloud fields. The aforementioned parameters are only evaluated for deep convection, but organization might function differently for shallow convection.

To inform studies of convective cloud-climate feedbacks and convective parametrization, in this study the available object-based organization parameters are assessed. Their performance is compared and an overview of their advantages and disadvantages is made. Additionally, an unsupervised machine learning approach to determine clusters in the cloud field is explored. The data consists of a large-domain ICON-LEM simulation (ICOsahydral Nonhydrostatic Large Eddy Model, (Dipankar et al., 2015)) of Trade wind cumulus with resolutions high enough to resolve small scale clouds. The great number of resolved clouds present in the domain provides a statistical significance that is required for a robust diagnosis of spatial organization in cloud populations.

After a description of the model data and the method that is used to determine the size and location of the clouds, the organization parameters that are compared are discussed in detail. We then show the results for the different organization parameters which is followed by a discussion about their respective advantages and disadvantages. A final analysis and summary conclude the study.

5.2 Data and cloud clustering algorithm

The details of the data used in this study can be found in Chapter 4, for which the same data set is employed. A large-domain LES simulation is used, performed with the ICON model (Zängl et al., 2014) and a set-up similar to Klocke et al. (2017). The inner domain of in total four nests is used, it measures about 150x400 km with a horizontal resolution of 150 m. It is situated over the subtropical Atlantic, directly east of Barbados, a location known for the regular occurrence of shallow cumulus clouds (Nuijens et al., 2014; Medeiros and Nuijens, 2016). The total duration of the simulation is 12 hours, all these 12 hours are included in the analysis. This is in contract with the previous chapter, where spin-up time was excluded. In this study it is included because organization is very pronounced during the spin-up period, it serves to highlight the differences between the compared organization parameters. Many clouds (up to 4500) are present in the 3D cloud fields which is significantly higher compared to normal-domain LES (Neggers et al., 2003).

In this study the vertically projected two-dimensional cloud field is used for the analysis. The details for the detection of the clouds, their sizes and locations, are described in detail in the previous chapter. Adjacent grid cells with a positive liquid water content are considered to belong to one cloud. The cloud size is defined as the radius of a circle with the same area as the cloud (Rieck et al., 2014). The location of a cloud is the center of the circle. All distances between clouds in this study are great circle distances, they are the distance from cloud center to cloud center corrected by the curvature of the earth.

5.3 Organization parameters

Organization is here defined as a clustered appearance of clouds. If the spatial distribution of clouds is significantly different from a homogeneously random distribution or regularly distributed clouds, the cloud field is characterized as organized. The parameters available to quantify the degree of organization are termed organization parameters.

5.3.1 Organization Index

Early examples of studying the spatial distribution of clouds and its randomness are presented by Zhu et al. (1992) and Nair et al. (1998). They applied a method which is based on the nearest neighbour spacing. The cumulative distribution of the nearest neighbour distances (nncdf) can be directly compared to a field with randomly distributed points, following the Poisson distribution (Fig. 5.1). By plotting the random nncdf (x-axis) against the simulated nncdf (y-axis), the difference between the two can be assessed. If the resulting line lies on or around the diagonal, the distribution in the field is random. If the line lies above the diagonal it means that distances in the simulated field are smaller than in the random field and that the clouds are therefore more organized. A regular distribution would result in a line below the diagonal. A figure like figure 5.1 can be summarized by integration. An integrated value greater than 0.5 means a clustered field, lower than 0.5

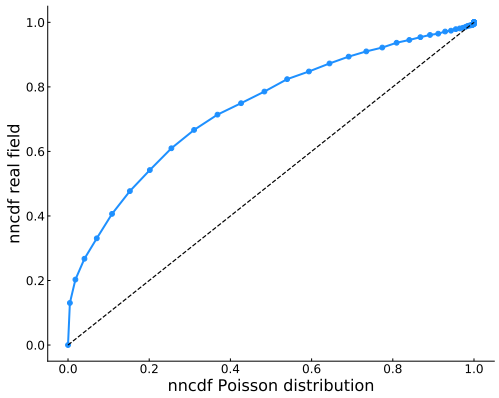

Figure 5.1: An example of a cumulative distribution of nearest neighbour distances (nncdf) figure, used to compute the I_{org}. On the x-axis the nncdf for a Poisson distribution which represents randomness and on the y-axis the nncdf for a simulated field. The blue line lies above the diagonal (black-dashed line), indicating that the clouds in the field are more organized than random. The integrated value of the blue line is the I_{org}.

means a regular field. The integrated value is referred to by Tompkins and Semie (2017) as the Organization Index (I_{org}). A caveat of this method is that there is the risk of misinterpreting I_{org}, as a combination of clustering and regularity can cancel each other out. Then the resulting I_{org} might indicate a random field where in reality both clustering and regularity are present, only on different scales.

5.3.2 SCAI and COP

Two other parameters that are developed for assessing the degree of organization in the field are the Simple Convecive Aggregation Index (SCAI) (Tobin et al., 2012) and the Convection Organization Potential (COP) (White et al., 2018). They are both based on the total number of clouds and the average distance between them. SCAI is defined as:

$$SCAI = \frac{N}{N_{max}} \frac{D_0}{L} * 1000, \tag{5.1}$$

with N the number of clouds, N_{max} the maximum number of clouds possible in the field (the total number of grid cells divided by two) and D_0 the geometrical average of the distances d for all pairs of clouds. L is the characteristic length scale of the domain. It is not clearly defined what this characteristic length scale should be, we will use the length of the diagonal of the domain (as this is the maximum distance). SCAI can be interpreted as the degree of disaggregation, normalized by the potential maximum disaggregation.

As presented by White et al. (2018), COP can differentiate between scenes that give the same SCAI value. This is because cloud size is part of the parameter as well, it is used to define a dimensionless interaction potential $V(i, j)$:

$$V(i,j) = \frac{\sqrt{A_i} + \sqrt{A_j}}{d(i,j)\sqrt{\pi}}, \tag{5.2}$$

where A is the area covered by the cloud and d is the distance. All pair-wise interaction potentials are summed and then divided by the total number of pairs in the field:

$$COP = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} V(i,j)}{\frac{1}{2}N(N-1)}. \tag{5.3}$$

COP increases for larger clouds and smaller distances and decreases for smaller clouds and bigger distances.

Both SCAI and COP lack a benchmark of randomness, the value of the parameter itself only indicates the amount of organization relative to other fields. For a better comparison with the other organization parameters, a field with randomly distributed clouds is used to normalize SCAI and COP. This field is computed by distributing the amount of clouds from the simulation randomly in the domain. Since for COP cloud size has to be taken into account as well, the sizes of the randomly distributed points are sampled from the cloud size distribution from the simulated field. Treating clouds as points and not as objects with a finite size when distributing them in the domain causes most likely some overlap of clouds when assigning a size to them. Some testing with fields with different amounts of overlap showed little difference for the values of SCAI and COP, we therefore assume that this effect is too small to influence the results because of the large amount of clouds. For every random field SCAI and COP are determined and this is repeated 10 times per timestep. The resulting random SCAI and COP values are the average of this 10 values. For normalization the values of SCAI and COP for the simulated field are divided by their respective values for the random field, hence a value of 1 means that the simulated field is close to random. Note that for plotting the value of SCAI is flipped, meaning that higher values mean more organization, just as for the other organization parameters.

5.3.3 Radial Distribution Function

Another way of assessing the relative cloud organization is using a radial distribution function (RDF). The RDF essentially is the cloud number density as a function of distance from the cloud. The result is compared to a random Poisson distribution. One of the firsts applying this method were Nair et al. (1998), later followed Cohen and Craig (2006) and Rasp et al. (2018). These three studies all found clustering at small distances from the clouds, and random distributions at larger distances. For the computation of the RDF circles with increasing radii are drawn around every cloud. Per circle the cloud number density is determined. The results are averaged over all clouds. The normalized RDF is defined and implemented as described in the jupyter notebook of Rasp et al. (2018):

$$RDF(r) = \frac{\left\langle N(r \pm 0.5\Delta r) \right\rangle}{A(r \pm 0.5\Delta r)} \frac{1}{\rho}, \tag{5.4}$$

where N is the number of objects, r the radii of the circles around the clouds and the angled brackets represent the mean over all objects. A is the area covered by the circles of radius r and ρ is the object number density over the whole domain. The RDF is unitless and tells how much more likely than random it is to find an object within distance r to another object. Without clustering in the field, the RDF would give a value of 1 for all r, this is the value for a random Poisson distribution.

5.3.4 Hierarchical clustering

As an alternative to the previously discussed organization parameters we apply an unsupervised learning method, referred to as hierarchical clustering, or hierarchical cluster

analysis (Hastie et al., 2009). This method can be applied to a large variety of data and is applied to find groups of objects that are similar to each other, but different to objects from other groups. Applying hierarchical clustering to atmospheric data is a novel approach, but from the data science perspective it is a natural approach to the problem of quantifying clustering. Hierarchical clustering has the advantage of not having to define the number of clusters beforehand, unlike other clustering methods. The only input hierarchical clustering requires is a measure of the dissimilarity between the objects.

For the current application the dissimilarity between the clouds is their distance. For the grouping of clouds into clusters a bottom-up approach is used, meaning that every cloud starts as its own cluster. For the clustering itself are several methods available, the results shown here are acquired by using the 'centroid' method or group average linkage (Hastie et al., 2009), which works as follows. First, the two clouds which have the smallest distance between them will be grouped into one cluster. Second, the center of this new cluster will be used to calculate distances to all other clouds again. These two steps will be repeated until all clouds are part of one cluster. The python library SciPy (Jones et al., 2001) is used to perform the clustering.

The result of the hierarchical clustering can be visualized and analysed in several ways. One typical outcome is a dendrogram, of which an example is shown in Fig. 5.2. This dendrogram is computed for only a small part of the complete domain (20x17 km) with 38 clouds. A dendrogram of the full domain would be more difficult to interpret, therefore a smaller and simpler version for a small subdomain is shown as an example. The individual clouds all start as a separate cluster on the x-axis. When the distance between the clusters increases (y-axis) they are merged into clusters consisting of one or more objects, this merging is indicated by the horizontal lines. The vertical lines show which clusters are merged and their length gives the distance at which this happened. This distance between clusters can serve as an indication of the amount of clusters present in the field. A relatively large distance between two clusters could indicate that they are actually two separate clusters. In the dendrogram in Fig. 5.2 the distance of 6 km is indicated as the cut-off by a black dashed line. This cut-off distance would mean that there are three clusters present. For the analysis in this study the largest increase in merging distance is derived from the dendrogram and is used to determine the amount of clusters in the cloud field. This method of finding the amount of clusters is very straightforward, but it does have its drawbacks. It can give for example the wrong amount of clusters if the variance in distance varies considerably among the clusters present. However, we will assume that this does not play a big role in our dataset.

5.4 Results

5.4.1 Cloud field characteristics

The temporal evolution of the cloud field is illustrated with timeseries of cloud cover, total number of clouds and maximum cloud size (Figure 5.4). Snapshots show the spatial distribution for three timesteps (Figure 5.3), indicated with coloured dots in the timeseries.

In the spin-up phase of the simulation large clouds are present. The dimension of these clouds stems from the coarser resolution of the domain that forces the inner domain. The large clouds should not considered to be very realistic, but they are included in the analysis to explore the behaviour of the organization parameters across a parameter space of cloud sizes that is as wide as possible. The large clouds are not numerous, but do cause a relatively high cloud cover and feature clouds up to 25 km. The number of clouds keeps

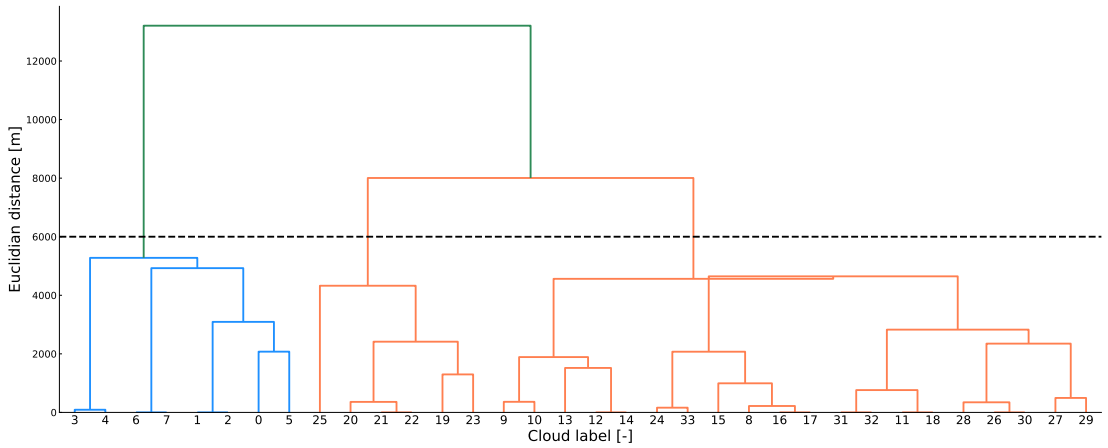

Figure 5.2: Example of a dendrogram for 38 clouds in a small subdomain. The individual clouds are presented on the x-axis at the bottom as individual clusters. A horizontal line indicates the merging of two clusters, the length of the vertical line indicates the distance between the two merged clusters (y-axis). The black dashed line shows an example of a cut-off distance, at this distance three clusters of clouds are present in the data.

increasing during the simulation, an indication that the large structures of the spin-up phase are falling apart. The drop in cloud cover and maximum cloud size at the beginning of the simulation support that the large clouds at the start rain out or split to form more, but smaller clouds. The second half of the simulation is characterized by a slow increase of cloud cover and cloud number, whereas the maximum cloud size stays constant around 5 km, except from a peak around 18:00 UTC. The transition from few big clouds to more smaller clouds is also visible in Figure 5.3. In these snapshots the cloud organization can be observed as well. The organized, large scale structures of 13:50 UTC change for smaller clouds that are either randomly distributed in the domain or form cloud streets.

The vertical structure of liquid water is visualized with profiles of cloud cover and cloud water averaged over the domain (Figure 5.5). Vogel et al. (2016) show similar vertical profiles for the same cloud regime. The spin-up phase is characterized by high cloud covers and high cloud water amounts at a height slightly above 2 km. These high cloud water amounts might be originating from the nocturnal boundary layer, since the simulation starts in the early morning local time. The high cloud fractions at this height might be due to outflow just below the inversion height. As time progresses these high clouds rain out and the low-level clouds become more pronounced, although their liquid water content is lower than for the high-level clouds. The decrease of projected cloud cover over time (Fig. 5.4a) can be explained by the disappearance of the high-level clouds. The increase in cloud number (Fig. 5.4b) mainly happens at lower levels, since there the cloud cover increases.

In studies on convective self-aggregation a reduction of high clouds is found for aggregated simulations (Wing and Cronin, 2015). Here we notice larger amounts of cloud water at high levels for the more organized cloud field, this might indicate that the organization we see here has indeed its origin in the coarser resolution from the bigger domains as opposed to the mechanisms for self-aggregation described in literature (Wing and Emanuel, 2014). The organization observed in our cloud fields thus has a different nature, but the large variation in cloud sizes helps to fully explore the limits of the parameters that are studied.

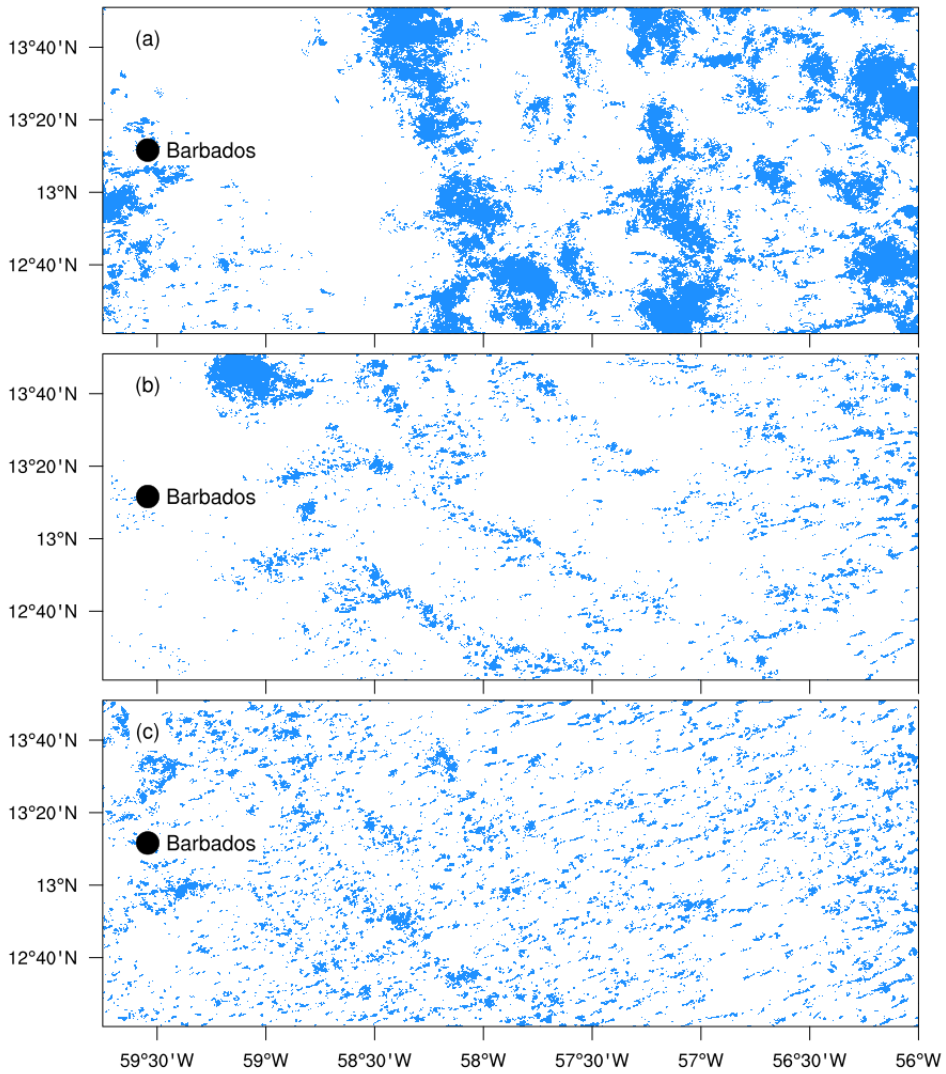

Figure 5.3: Three snapshots of projected cloud liquid water in blue, for a) 13:50 UTC, b) 17:50 UTC and c) 22:50 UTC.

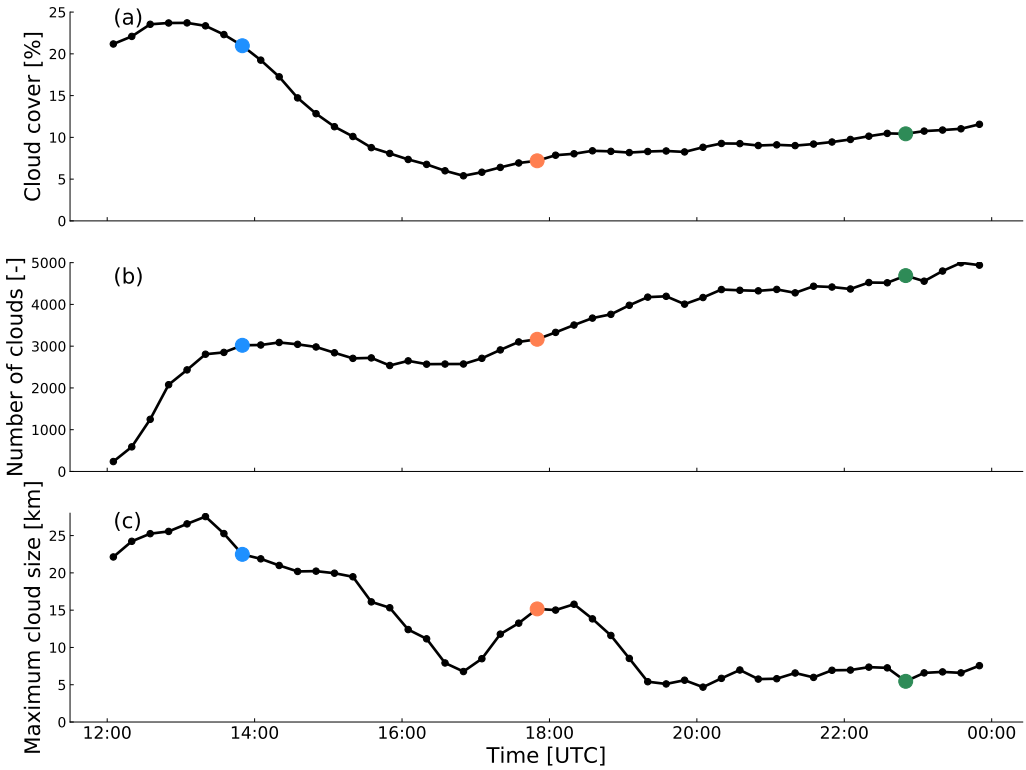

Figure 5.4: Timeseries for a) projected cloud cover, b) total number of clouds, and c) the maximum cloud size. The coloured timesteps refer to the snapshots in Figure 5.3.

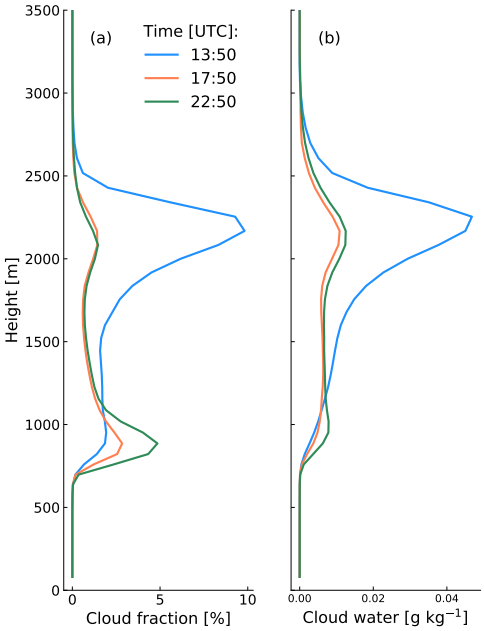

Figure 5.5: Vertical profiles of a) cloud fraction and b) cloud water averaged over the domain. The three different profiles correspond to the snapshots of Figure 5.3 and the coloured timesteps in Figure 5.4.

5.4.2 Organization parameters I_{org}, SCAI and COP

The time evolution of I_{org}, SCAI and COP is shown in Figure 5.6a, b, c. Values at the black dashed line indicate randomness, therefore of these three parameters only I_{org} shows strong organization at the beginning of the simulation. This is not the case for SCAI and COP, both of them have values below the reference level of randomness at the simulation start. SCAI continues to increase rapidly and shows a relatively high level of organization for the period between 14:00 and 17:00 UTC, just like I_{org}. Both I_{org} and SCAI then decrease during the rest of the simulation, ending with values close to random. So even though I_{org} only uses nearest neighbour distances as opposed to SCAI which uses the average of all pair-wise distances, they both show a similar signal. Earlier studies suggest that SCAI is mainly dominated by the number of clouds and not by distances between them (White et al., 2018). Comparing SCAI to the timeseries of the number of clouds shows that this is not the case here. However, because in our study we divided SCAI for the real field by a SCAI for a random field with the same number of clouds, the effect of the number of clouds is not visible any more. Also in our studies the time evolution of the non-normalized SCAI is very similar to the time evolution of the number of clouds (not shown).

COP shows a somewhat different pattern than I_{org} and SCAI. At the start of the simulation, COP values below the reference level of randomness (meaning regularly distributed clouds) are seen, although COP increases later on. Even so, COP stays close to random until the end of the simulation. If anything, it increases slightly at the end. This increase is opposite behaviour to I_{org} and SCAI. A reason why COP does not show a transition from more organized to less organized cloud fields could be the fact that it includes the cloud size. Following the definition of COP, it increases for larger clouds and smaller distances and decreases for smaller clouds and bigger distances. However, the results from chapter 4 show an increasing distance for increasing cloud size. That would mean that including the cloud size the way it is done for COP balances out the effect increasing or decreasing distances would have. The compensating behaviour between cloud distance and cloud size could explain that COP does not give a strong signal on the degree of organization. More research is needed to give a definite explanation.

5.4.3 RDF

Figure 5.7 shows the RDF for the three timesteps of Figure 5.3 as a function of radius. The steps with which the radius increases have a width that is the same as the grid resolution. For all three timesteps the RDF decreases with distance from the cloud, indicating that the cloud number density closer to the cloud is higher than can be expected for a random field. The shape of the RDF is similar for the three timesteps, although values for 13:50 and 17:50 UTC are higher than for 22:50 UTC. These higher values mean a higher degree of clustering, at least for a large range of radii. At large radii the difference is minimal. Since the shape for all timesteps is similar and only the magnitude of the RDF values differs, only the maximum value of every timestep is included in the timeseries of Figure 5.6. The maximum value of the RDF of the third timestep is extremely high (about 75), therefore only the values from the fifth timestep onwards are shown. Over the course of the simulation the maximum RDF value shows a steady decrease, a similar trend as for I_{org} and SCAI. The value of one, for randomness, does not occur, meaning that according to the RDF the cloud fields show organization at all times.

Since the cloud fields contain a large amount of clouds, we will take a small sidestep and make use of the data at our disposal to look for any relation between cloud size and degree of organization. For this, the data is sorted on cloud size into a histogram, with the bin

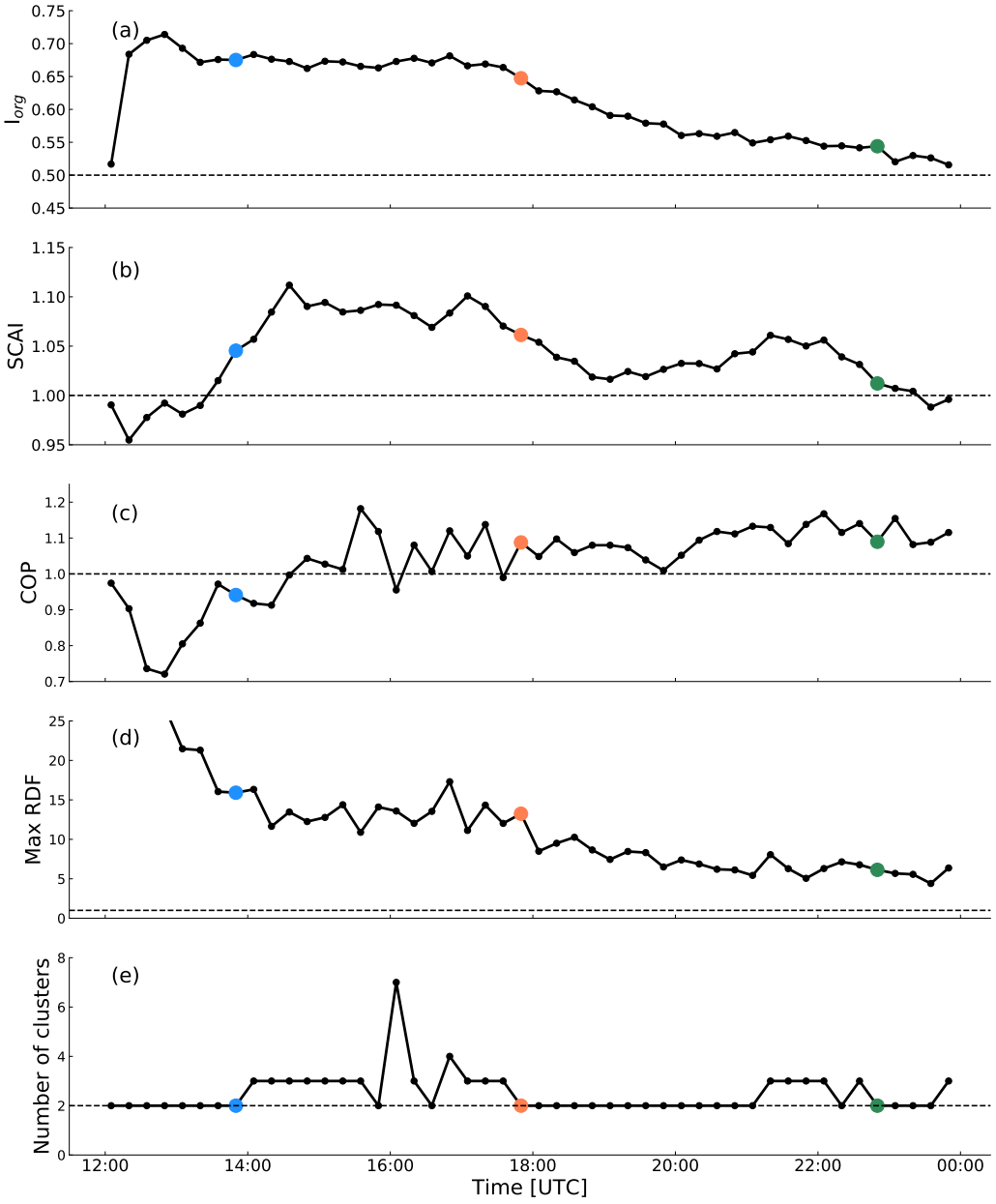

Figure 5.6: Time evolution of I_{org}, SCAI, COP, the maximum RDF value and the number of clusters based on hierarchical clustering. The horizontal black dashed line indicates the value for a random spatial distribution. The coloured timesteps refer to the snapshots in Figure 5.3.

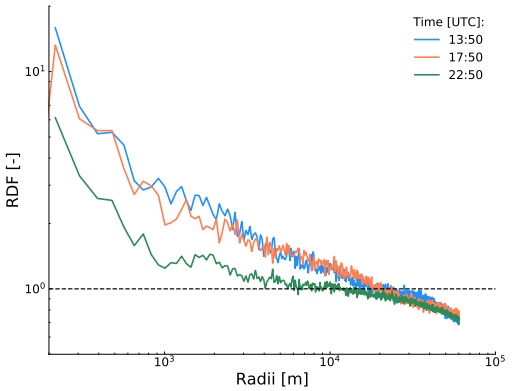

Figure 5.7: Normalized radial distribution function for the three cloud fields shown in Figure 5.3. The RDF gives how much more likely than random (dotted line) it is to find a cloud as a function of distance (radius) from the cloud.

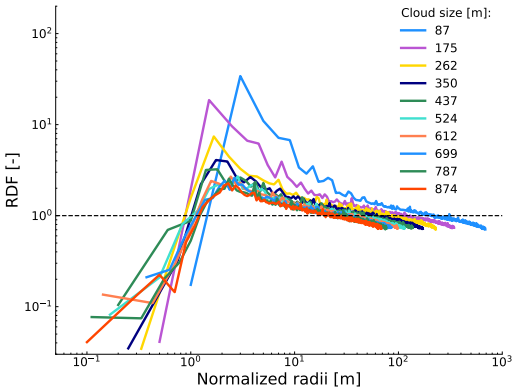

Figure 5.8: RDF values averaged per cloud size. Every colour depicts the RDF for a certain cloud size, only the 10 smallest cloud sizes are shown. The values in the legend are the upper edges of the bins for cloud size.

width the resolution of the simulation. The RDF averaged per cloud size for the first 10 cloud size bins is shown in Figure 5.8. The radii at the x-axis per cloud size are divided by the radius at which the maximum RDF occurs. This figure shows two things. First, the smaller clouds show more organization than the larger clouds, this is in accordance with the findings of chapter 4 and also corroborates what is found in Neggers et al. (2019) on the internal variability of the cloud size distribution. Second, the peaks of the RDF align pretty well, except for the peak for the smallest cloud sizes. The alignment at a normalized radius of about 20 m happens because we normalized the radii with the radius of the peak RDF value. The alignment of the peaks indicates a positive relation between the distance at which the strongest organization happens and the cloud size. This is easily explained by the way the RDF is defined: it uses cloud center distances. Cloud center distances are by definition bigger for larger clouds since their radius is bigger. Therefore, the bigger the cloud, the larger the minimal distance at which other clouds can occur. To check if this is indeed the underlying reason for the relation between cloud size and distance of maximum organization, the RDF per cloud size is computed using cloud edge distances instead of cloud center distances (not shown). The cloud edge distance is the cloud center distance minus the radii of the two clouds under consideration. When correcting for the radii of the clouds, the relation between cloud size and maximum organization distance is not present any more. For all cloud sizes the highest degree of organization happens at the smallest distance and there is no effect of cloud size.

5.4.4 Hierarchical clustering

The results of using hierarchical clustering to determine the number of clusters per timestep are shown in Figure 5.6e. For most cloud fields the automatic detection of the number of clusters gives a value of two. Slightly more clusters are found for fields around 16:00 and 22:00 UTC. This could indicate that there is more clustering happening around these hours, which causes more clearly defined clusters. However, the large amount of cloud fields where just two clusters are present is not very encouraging. Using the largest increase in merging distance as a threshold for the amount of clusters namely gives a minimum of two clusters. It could therefore very well be that there are no clearly defined clusters present at these times and that the higher values are not very meaningful either.

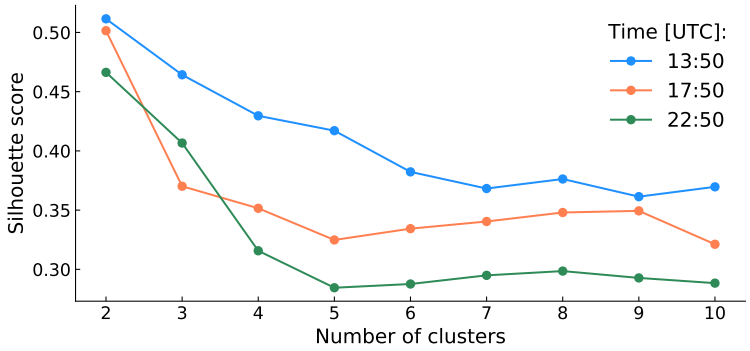

Figure 5.9: The silhouette score for three timesteps. Clustering of a high quality and therefore clearly defined clusters would give a peak in the silhouette score.

Since using the largest increase of merging distance as threshold for the amount of clusters did not yield convincing results, also the silhouette coefficients are determined. The silhouette coefficient gives for each cloud information of how close it is to its own cluster and how far away from the other clusters. (Rousseeuw, 1987). For the silhouette coefficient first a number of clusters has to be set. With the help of the dendrogram, the clouds belonging to every cluster are determined. It is defined as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \tag{5.5}$$

with b the average distance of cloud i to all clouds in the other clusters. a is the average distance between cloud i and all other clouds in the same cluster. The values for s range from -1 to 1. The higher s, the closer a cloud is to its own cluster and the further away from the neighbouring clusters. To obtain a single value for a cloud field with a certain number of clusters, the silhouette coefficients for all clouds are averaged, yielding the *silhouette score*. It gives information on the quality of the clustering of the data, a high quality means a clear separation of clusters. The silhouette score is computed for every field and for a range of number of clusters.

For the three example cloud fields of Figure 5.3 the silhouette scores for a number of clusters ranging from two to ten are shown in Figure 5.9. For the other timesteps a very similar signal is found. If the data would have a clearly defined number of clouds, the silhouette score should peak at that value. With a highest score for two clusters and only decreasing scores for higher numbers of clusters, there is no such clear peak present for the cloud fields. In general, with maximum values around 0.5, the silhouette scores are fairly low. These findings corroborate the analysis using the largest increase in merging distance: the quality of clustering is low.

Even though no clearly separated clusters of clouds can be found, still the bottom-up approach of clustering clouds can provide useful information. As the clouds that belong to a certain cluster are known, the properties of the different clusters can be studied. It can be seen as a more advanced way of subsampling clouds from a big domain. Examples of cluster properties are: the size, the number of clouds, the average distance between the clouds, and the cloud size density (CSD). As an example we will look into the normalized CSD of the clusters of the cloud field at 16:05 UTC, this is the time where the peak of seven clusters occurs, according to the hierarchical clustering. The distribution of these clusters over the domain shows that the amount of clouds per cluster varies significantly (it ranges from 195 to 616) (Fig. 5.10). The accompanying CSD for each cluster and for all clusters

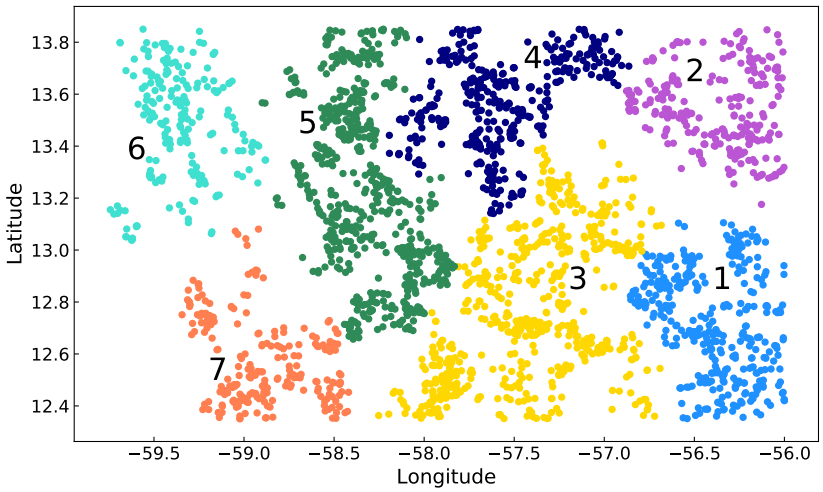

Figure 5.10: Spatial distribution of clusters of clouds according to hierarchical clustering. Every point is a cloud, and clouds with the same colour belong to the same cluster. Size is not taken into account in this figure.

together is shown in Figure 5.11. The CSD of the two biggest clusters (3 and 5, green and yellow) show similarity with the CSD for the total field up to large cloud sizes, whereas the CSDs for the smaller clusters start to deviate at smaller cloud sizes. Nevertheless, the largest clouds are present in the smallest clusters (1, 6 and 7). In general do the CSDs of the clusters show that they only for the larger cloud sizes deviate significantly from the CSD for the total field.

According to Rasp et al. (2018), it also would have been possible to estimate the average size of the clusters by using the RDF. They state that the distance from the cloud at which the RDF reaches values lower than 1 can be seen as the radius of the cluster. However, using twice this radius as the diameter of a cluster as cut-off distance for the dendrogram (any merging distance larger than this would mean a different cluster) results on average in more than 300 clusters. These clusters are hard to analyse since they consist of a small amount of clouds and show a high variability.

Table 5.1: Overview of how the different organization parameters perform for several criteria. The computing time only serves as an example, it is determined for a field with about 4500 clouds.
*not shown in literature, but normalizing SCAI as we do in this study might give good results.

	I_{org}	SCAI	COP	RDF
Shows organization transition	yes	yes	no	yes
No relative measure	yes	no	no	yes
No visual inspection needed	no	yes	yes	no
Works for deep convection	yes	no*	yes	yes
Computing time (CPU time) [s]	34.3	**32.9**	36.4	35.7

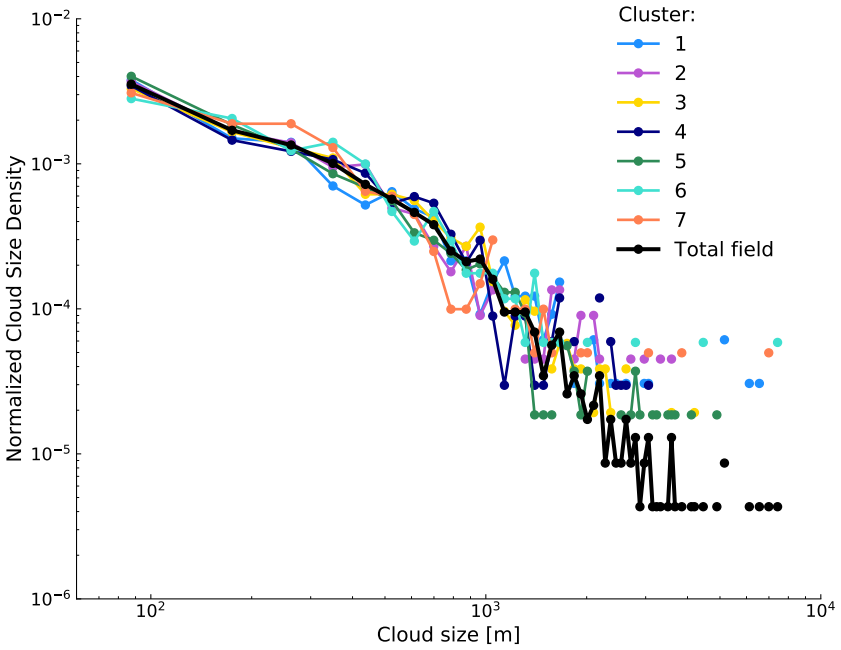

Figure 5.11: The normalized cloud size density for all clusters from figure 5.10 (coloured) and for the total field (black).

5.5 Discussion and Conclusions

In this study several parameters for assessing the level of organization have been applied to a field of shallow cumulus clouds. Every organization parameter has its advantages and disadvantages, but most of them showed a transition from more organized to less organized cloud populations during this particular simulation of Trade wind cumulus. This transition can also be detected by eye from snapshots. I_{org}, SCAI and RDF apparently capture the behavior in spatial organization that most humans can detect by eye. COP does not, and hierarchical clustering suggests that no clearly separated clusters can be distinguished.

Earlier studies have applied these parameters mainly to deep convection, with less, but bigger clouds. White et al. (2018) show that SCAI is heavily influenced by the number of clouds in these cases, and therefore they developed COP. With the added information in the form of cloud size COP was able to distinguish cloud fields with a similar amount of clouds, but different degrees of organization. However, when normalizing SCAI with a random field, as we do in this study, the number of clouds does not determine the value of SCAI and COP has no added value. I_{org} is an organization parameter that has been applied to deep convection as well. Tompkins and Semie (2017) opted for this parameter since it is not a relative measure (like SCAI and COP), but allows direct classification of a cloud field as regular, random or clustered. In their study it served well to quantify organization, although it does not make a distinction in scales of convection. Being able to separate scales should be an aspect of a meaningful organization parameter according to Wing (2019). Summarizing, I_{org}, SCAI and RDF perform for our shallow cumulus simulations, where I_{org} and RDF have been shown to perform for deep convection as well.

Putting all the advantages and disadvantages together (Table 5.1), we see that there is no organization parameter that clearly performs best. In this study only COP does not give satisfactorily results since it does not show the transition from more to less organized cloud fields. Also is COP, together with SCAI, a relative measure, making it difficult to interpret the output, unless a comparison is made to a random field as is done here. Note that

then the computing time at least doubles, since also the value for a random field has to be computed. As described in the previous paragraph, SCAI does not perform well for a small number of clouds and would therefore not be the ideal candidate for usage in a scale-aware parametrization scheme. For both I_{org} and RDF it is necessary to visually inspect the result before the single value output can be trusted. If one of these two parameters would be implemented in a convection scheme this inspection has to be done automatically.

For ongoing efforts to make convective parametrizations capture effects of spatial organization it is informative to compare the computational cost of all parameters tested in this study. To give an indication of computing costs for the studied parameters, an example of the CPU time is included in Table 5.1. It is determined for the calculation of the value for one specific field on a single CPU, this is the field at 22:50 UTC with about 4500 clouds. Note that the computing time behaves non-linear with the amount of clouds, since for all parameters the pair-wise cloud distances have to be computed. The coding and computer specifications influences the computing time, the numbers in the table therefore merely serve as an example and to be able to compare the different parameters. According to the resulting computing times, SCAI is fastest and COP takes longest, although the spread in the values is small. I_{org} is the second fastest to compute and also scores well in general, being able to perform over a range of scales. It depends on the nearest neighbor spacing, a parameter which can be useful for developing stochastic and scale-aware convective parametrization schemes (Chapter 4 and Neggers et al. (2019)).

Object-based organization parameters are not the only way to assess organization of convection. An interesting new approach is the recently applied classification of cloud populations into 'fishes', 'flowers', 'gravel' and 'sugar' [1]. The performance of the here presented parameters on these classifications would be interesting to see. Hierarchical clustering might give interesting results, especially for the 'fishes' and 'flowers'. Some adjustments might be needed for it to perform well, so could one switch to use single linkage instead of group average linkage. The assumption of circular objects does not hold for these large structures, so cloud edge distance instead of cloud center distance could help in improving results for these type of structures. Another type of regularly encountered organization is squall lines, or cloud streets. For these features to be recognized by a parameter, also cloud edge distance might be advisable. Another not-object based method that could capture organization in lines is the wavelet based WOI method (Brune et al., 2018) since it gives additional information on the orientation of the objects. None of the parameters presented in this study gave a clear sign of recognizing the street like organization that is visible in some parts of the snapshots in Fig. 5.3. A disadvantage of this method is that it uses precipitation as input parameter and it is therefore not well applicable to non- or weakly precipitating clouds like fair weather cumulus. An optional new parameter might be inspired by the area of applied topology, where recent advances have shown that the technique of persistent homology can be used as a multiscale parameter for spatial data which shows cluster patterns (Robins and Turner, 2016; Hiraoka et al., 2016).

The data used in this study comes from a large-domain simulation performed over the ocean, a relatively homogeneous surface. The long time series of present clouds make it possible to compare the values for the organization parameters and draw conclusions based on that. However, surface heterogeneity and the diurnal cycle might play an additional role in convective organization, this is explored in the next chapter. For increasing the parameter space to which the organization parameters are applied, one can think of using a data set of many simulated shallow cumulus days, as is done in Chapter 3, or a composite of many shallow cumulus days (Zhang et al., 2017).

[1] https://raspstephan.github.io/research/sugar-flower-fish-gravel/

Chapter 6

Surface heterogeneity in nested simulations and the effect on cumulus organization

Abstract

The formation of shallow cumulus clouds depends strongly on the surface conditions. In this study, we use an ensemble of Large Eddy Simulations with different surface properties to assess the influence of the surface on spatial patterns of cumulus cloud fields. Two days showing shallow cumulus clouds are selected. For these days, simulations are performed at a continental site in north western Europe. A sensitivity study includes both diminished and enhanced topography, as well as a modified distribution of land use types. These changes in the surface conditions lead to differences in the cloud field. For the quantification of the differences both cloud size distributions and the Organization Index (I_{org}) are applied. These two methods capture the surface effects on cloud sizes and on the spatial distribution of the clouds, respectively. Additionally, a new approach for quantifying spatial organization based on the vertical velocity is explored. The surface conditions do not affect the slope of the cloud size distribution, but the range of cloud sizes does change among the simulations. Generally, the simulations with enhanced topography show larger maximum cloud sizes. I_{org} shows no significant differences between the simulations and every cloud field is classified as clustered. However, modifying the land use distribution resulted in higher I_{org} values for a simulation where a cloud street is present. The spatial difference in heat fluxes create for this simulation a quasi-secondary circulation. Since I_{org} shows only small differences for simulations with different surface conditions, the variance of the vertical velocity is studied as a possible alternative or additional indicator of organization. It shows more differences between the simulations, which might suggest that there is indeed more variation than I_{org} suggests, but the absence of a clear benchmark complicates evaluation. In this study we show that with a realistic model set-up, the impact of topography and land use on the spatial distribution of shallow cumulus can be observed. Even though quantification of the impact is not straightforward, the results of this study provide new insights on the interaction between land surface and cloud formation.

6.1 Introduction

The spatial organization of convection is an active area of research (Wing, 2019). Studies have shown spontaneous organization in the absence of surface heterogeneities of which the underlying processes are not yet fully understood (Bretherton et al., 2005). However, heterogeneity in surface conditions has an influence on atmospheric flow and could therefore play a role in convective organization as well (Pielke, 2001). This is especially the case for shallow cumulus clouds, which are directly coupled to the surface and are therefore greatly affected by the surface conditions (Garcia-Carreras and Parker, 2011).

Usually, two types of surface heterogeneity are considered: static and dynamic heterogeneity. Static heterogeneity refers to surface conditions which vary little over time, e.g. soil moisture gradients (Cioni and Hohenegger, 2017). Dynamic heterogeneity is caused by external factors, an example is cloud shading. Lohou and Patton (2014) and Horn et al. (2015) both studied this phenomenon and found that fields of shallow cumulus clouds introduce dynamic heterogeneity, influencing the energy balance at the surface. However, this heterogeneity is not strong enough to create secondary circulations of significance, and it has been shown that these secondary circulations play an important role in the development of convection (Avissar and Schmidt, 1998; van Heerwaarden and de Arellano, 2008; Kang and Bryan, 2011; Huang and Margulis, 2013; Rieck et al., 2014).

Several modelling studies have shown that the development of secondary circulations in simulations with imposed surface heterogeneity creates preferential areas for clouds to form. Rieck et al. (2014) found that clouds usually organize over the warm and dry patches, where the moisture is transported to the dry patch via a secondary circulation. Lee et al. (2019) studied the dependence of secondary circulations on patch size and background wind speed. They concluded that low background wind speed and big patches promote strong secondary circulations where clouds form over the dry patches. Patton et al. (2005) concluded that the scale of the heterogeneity relative to the boundary layer height should be between four and nine for secondary circulations to develop. Apart from the patch size and the background wind speed, the amplitude of the heat flux difference between the patches also plays a role in the development of a secondary circulation (van Heerwaarden and de Arellano, 2008).

Strong and persistent secondary circulations control the air flow and thereby enhance organization. Besides surface heterogeneity being able to influence circulations, also topography is of importance. The presence of complex terrain influences turbulent and convective processes (Banta, 1984; Stein, 2004; Kirshbaum et al., 2018). Homar et al. (2003) describe how topographic features of different scales create small-scale circulations that can intensify convective storms.

Most of the above mentioned studies with imposed surface heterogeneity or topography make use of a highly idealized set-up. For the surface heterogeneity the different patches are organized in strips or follow a checker board pattern. Although this is useful for studying the processes influenced by surface heterogeneity, it does not reflect reality. A realistic representation of surface conditions becomes more and more important now that resolutions of weather prediction and climate models are getting higher. The impact of realistic surface conditions on cloud formation is a necessary process to understand.

We will therefore employ a regional-scale model with surface conditions originating from satellite observations. Thanks to ongoing model development and increasing computer power, we can use this model with an interactive land surface scheme at resolutions high enough to resolve shallow cumulus clouds. Utilizing this as realistic as possible model set-up, the surface conditions will be varied to study their influence on shallow cumulus

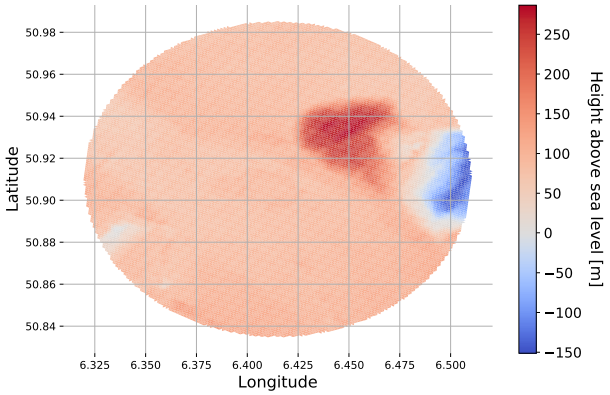

Figure 6.1: Topography for the control simulation for the analysed domain. The observational supersite JOYCE is located in the centre of the domain.

cloud formation, organization and the development of secondary circulations. Variables that will be altered are the amplitude of topography and land use, where land use includes several parameters like Leaf Area Index, roughness length and plant cover per grid cell. To quantify the effect of the varying surface conditions on the spatial distribution of shallow cumulus, the Organization Index (I_{org}) will be applied (for description, see Chapter 5 and Tompkins and Semie (2017)).

The details of the model set-up and sensitivity tests are given in the next section. A description of the two simulated days can be found in the section 6.3, as well as an assessment of the influence of the surface conditions. Based on two observed processes, an alternative to I_{org} is explored (section 6.4). A discussion and the conclusions are given in the last section.

6.2 ICON simulations

6.2.1 Model set-up

The simulations in this study are performed with the Icosahedral Nonhydrostatic (ICON) model (Zängl et al., 2014). Two days are selected and for both days the same set-up is used, consisting of four nested grids (Marke et al., 2018). The outer domain has a radius of 100 km, a horizontal resolution of about 600 m and is forced with IFS data. The inner domains are gradually getting smaller with a higher resolution. The most inner domain is used for the analysis, it has a radius of 10 km with 75 m horizontal resolution. The lowest 2 km of the domain contain 33 vertical levels with the smallest thickness 20 m. The temporal resolution of the output data is 10 minutes. The center of all domains is at JOYCE, an observational super-site in south-western Germany.

The surroundings of JOYCE show little variation in topography (Fig. 6.1). In the eastern part of the domain a hill is located, right next to it is a pit mine. In terms of land use there is some more variation. Input data for the land use is obtained from GLOBCOVER (Bontemps et al., 2012). This data is interpolated to our resolution settings. The land use data contains several variables, e.g. Leaf Area Index (LAI), roughness length and percentage of plant cover per grid cell. As an illustration of the land use, the Leaf Area Index (LAI) is shown in Figure 6.2. The majority of the domain consists of agriculture (orange), and the area with the pit mine is visible in the east. The simulations done with these surface conditions are defined as Control.

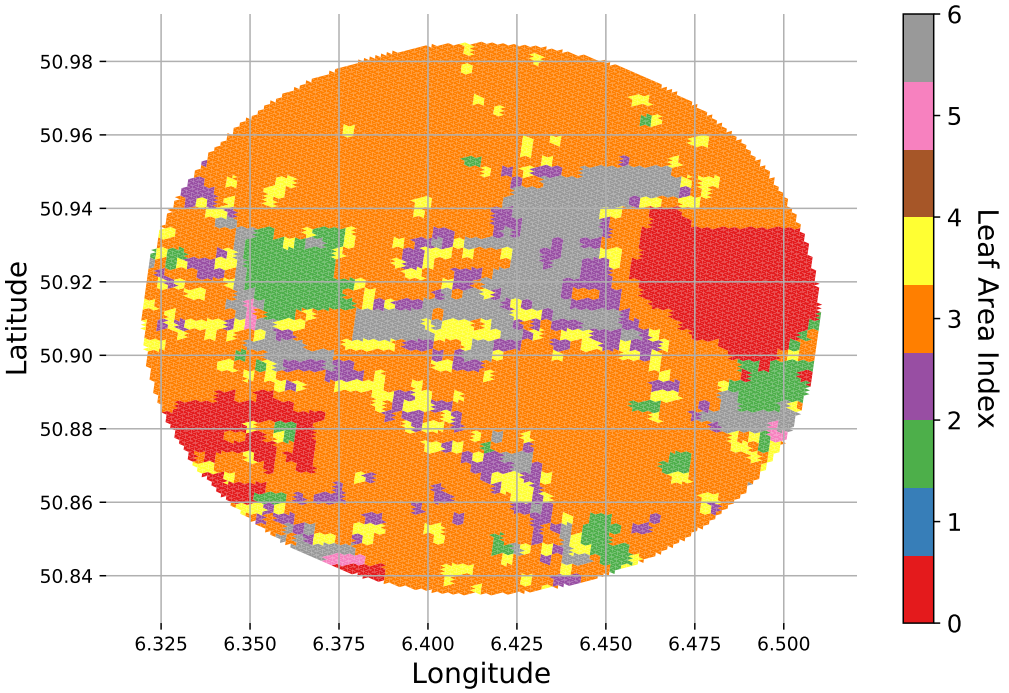

Figure 6.2: Leaf Area Index for the control simulation for the analysed domain.

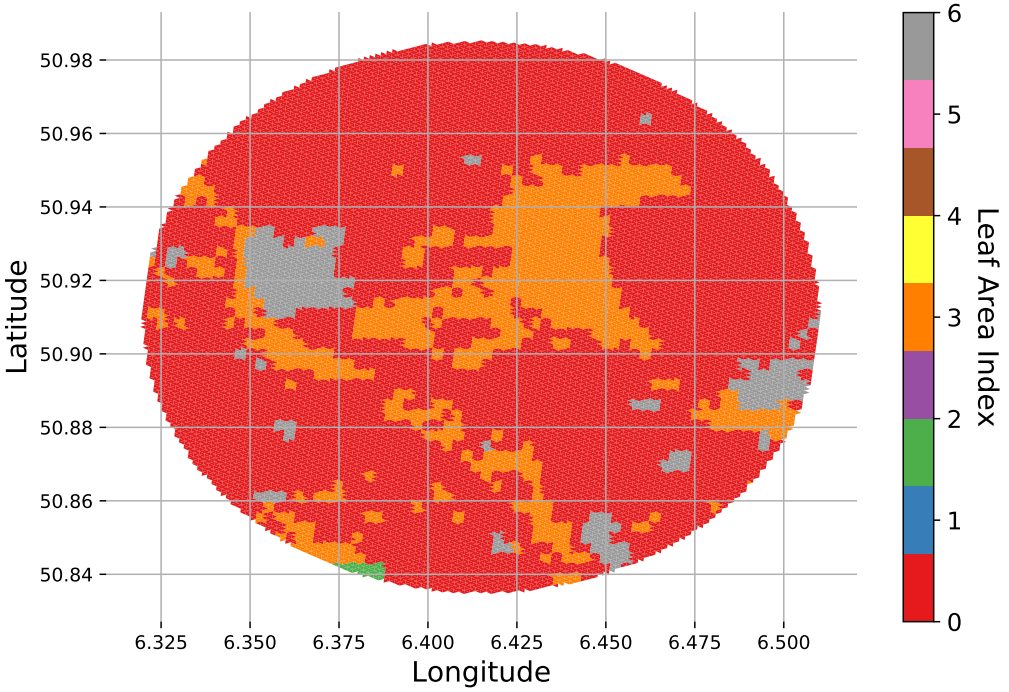

Figure 6.3: Leaf Area Index for LU for the analysed domain.

6.2.2 Simulated days

Two typical shallow cumulus days are simulated, 5 June 2013 (form now on called day A) and 25 July 2012 (from now on called day B). Based on measurements taken at JOYCE we can say that both days show a convective boundary layer that grows deeper during the day. For day A, there is rapid increase of boundary layer height in the morning, and shallow clouds start to form around 10:00 hr. After 2-3 hours the clouds disappear, the rest of the day shows mainly clear sky. For day B the growth of the boundary layer goes more gradually and clouds only start to form at 13:00 hr. For both days cloud fractions are low, with typical maximum values for shallow cumulus of 10-15 %.

6.2.3 Sensitivity tests

To test the sensitivity of cloud organization on surface conditions, two different types of experiments are performed. First, the topography is changed. For both days two extra simulations are performed. The amplitude of the topography in all four domains is multiplied by 2 (Topo2), and by 0.5 (Topo05). This makes the extreme topography stand out more and less, respectively. Second, the land use is adjusted (LU), following Marke et al. (2019). For this, the original data is divided into five land use type classes. These are crop/grassland, bare ground, urban, forest and water. These types are then shuffled around in the following way: crop/grassland to bare ground, bare ground to water, urban to forest, forest to crop/grassland and water to urban (Fig. 6.3). By changing the land use like this, the spatial distribution of sensible and latent heat flux extremes will be different, while at the same time the scale of the heterogeneity stays the same.

6.3 Results

In this section first the two Control simulations will be described. Then, we will go into more detail considering the cloud size distribution and the quantification of the degree of organization. The results section is ended with an analysis of the formation of a cloud street in one of the simulations.

6.3.1 Description of the two control simulations

The ICON simulations for Control for both days show typical shallow cumulus cloud characteristics. Day A is characterized by a well mixed boundary layer in terms of virtual potential temperature (θ_v) with slowly increasing surface temperatures over time (Fig. 6.4a). The inversion height increases from slightly below 2 km at 10:00 hr to about 2.5 km at 14:00 hr. The specific humidity (q_s) at the surface decreases after 10:00 hr (Fig. 6.4c), as moisture is transported upwards where it forms clouds. The temporal evolution of the cloud fraction is shown in Figure 6.5a. Clouds start to form shortly after 10:30 hr, and the cloud base slowly increases during the two subsequent hours. The depth of the cloud layer is fairly constant during that time period. Maximum cloud fractions reach 15 % and after 13:00 hr the clouds disappear.

The vertical profiles for day B are shown in Figure 6.4b,d. The surface temperatures are higher for this day than for day A, although the inversion height is lower. The profiles of θ_v do not change significantly over time, above a height of 1 km there is no difference for the three different timesteps. The q_s is also higher than for day A. Again the upward transport of moisture over time can be observed, with lowest surface values at 14:00 hr.

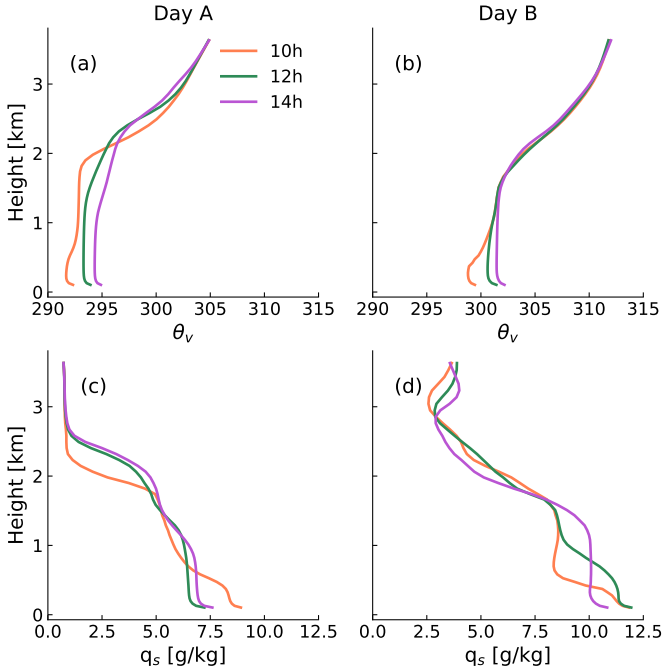

Figure 6.4: Vertical profiles for three times during Control for both days for virtual potential temperature (θ_v) in (a),(b) and specific humidity (q_s) in (c),(d).

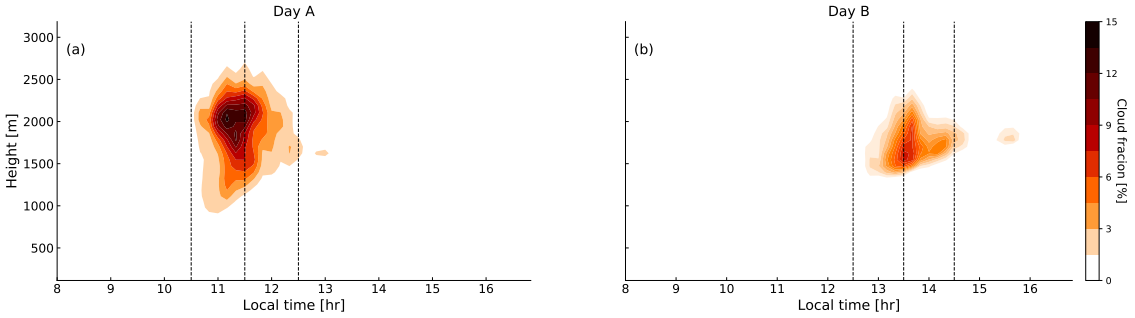

Figure 6.5: Cloud fractions for the control simulation of both days. The vertical dashed lines indicate the times used in the analysis for snapshots and averaging.

On day B, clouds start to form later compared to day A, despite the higher temperature and humidity values. At 13:00 hr clouds start to form, the cloud base increases gradually with time. Maximum cloud fractions of around 7 % are reached (Fig. 6.5b).

When looking at the surface fluxes for the two days, some interesting features can be observed (Fig. 6.6a). The difference in surface fluxes between the two days seems small, nevertheless the onset of cloud formation differs between these two days. The latent heat flux for day A increases faster than for day B and it reaches higher values. At the time of the highest cloud fractions (11:00-12:00 hr), a small dip is present, the same is true for the sensible heat flux. The timeseries for both the sensible and latent heat flux for day B show a typical sinusoidal form, the presence of clouds in the afternoon does not seem to influence this.

Day A shows a higher bowen ratio than day B for the first hours of the day (Fig. 6.6b). The higher bowen ratio stems from the higher sensible heat flux and lower latent heat flux during this time. During the day, the bowen ratio for day A decreases, with an extra dip

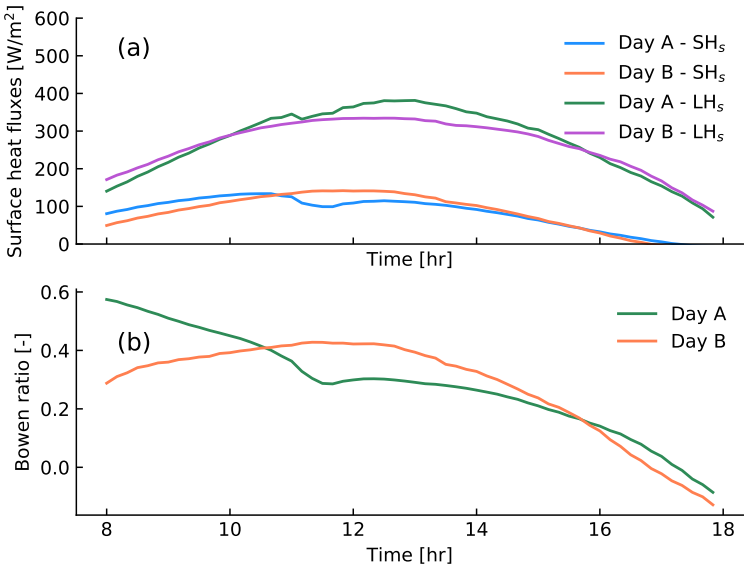

Figure 6.6: Diurnal evolution of the surface sensible and latent heat flux (a) and the bowen ratio at the surface (b) for both simulated days.

shortly before 12:00 hr. The bowen ratio for day B has a peak around 12:00 hr, after which it decreases and shows similar values as day A at the end of the day.

The spatial distribution of the clouds is shown with three snapshots of cloud cover per day (upper rows of Figure 6.7 and 6.8). Day A exhibits a large variation in the number of clouds over the short timespan of two hours. No clear influence from either topography or land use is present. The cloud cover for day B is lower, but clouds seem to form in a cloud street diagonally over the domain (north-west to south-east). There are no clouds present over the highest elevated area of the domain.

6.3.2 Spatial distribution of clouds

To directly demonstrate the effect of changing surface conditions on the spatial distribution of clouds, several snapshots of the cloud mask are shown in Figures 6.7 and 6.8, for day A and B respectively. The snapshots for Topo2 for day A show more and bigger clouds than for Control at 10:30 hr. An hour later, some scattered clouds are present, and at 12:30 hr there are some clouds in the middle of the domain. Hence, as time progresses, there is no clear increase and decrease of cloud number as for Control. Topo05 looks more like Control in terms of the development of the cloud field. At 10:30 hr there is only a small amount of clouds present, whereas an hour later almost the whole domain is covered in clouds. The area in the middle of the domain (around JOYCE) seems relatively clear of clouds. A small amount of clouds remain at 12:30 hr. This temporal evolution of the amount of clouds is similar for LU. At 11:30 hr, most clouds can be found at the edges of the domain, again the middle of the domain seems to be relatively cloud free.

For Topo2 on day B (Fig. 6.8) there seems to be a slight temporal delay in the formation of clouds. Most of the clouds are present at 14:30 hr as opposed to 13:30 hr for Control. At 14:30 hr, the clouds are mainly located over the region with the strongest topography. For Topo05, the clouds form in the same area as for Control, although there is a lower number of clouds present. For LU, the cloud street similar to the one seen for Control is already present at 12:30 hr. In this area more clouds are present at 13:30 hr, at that time

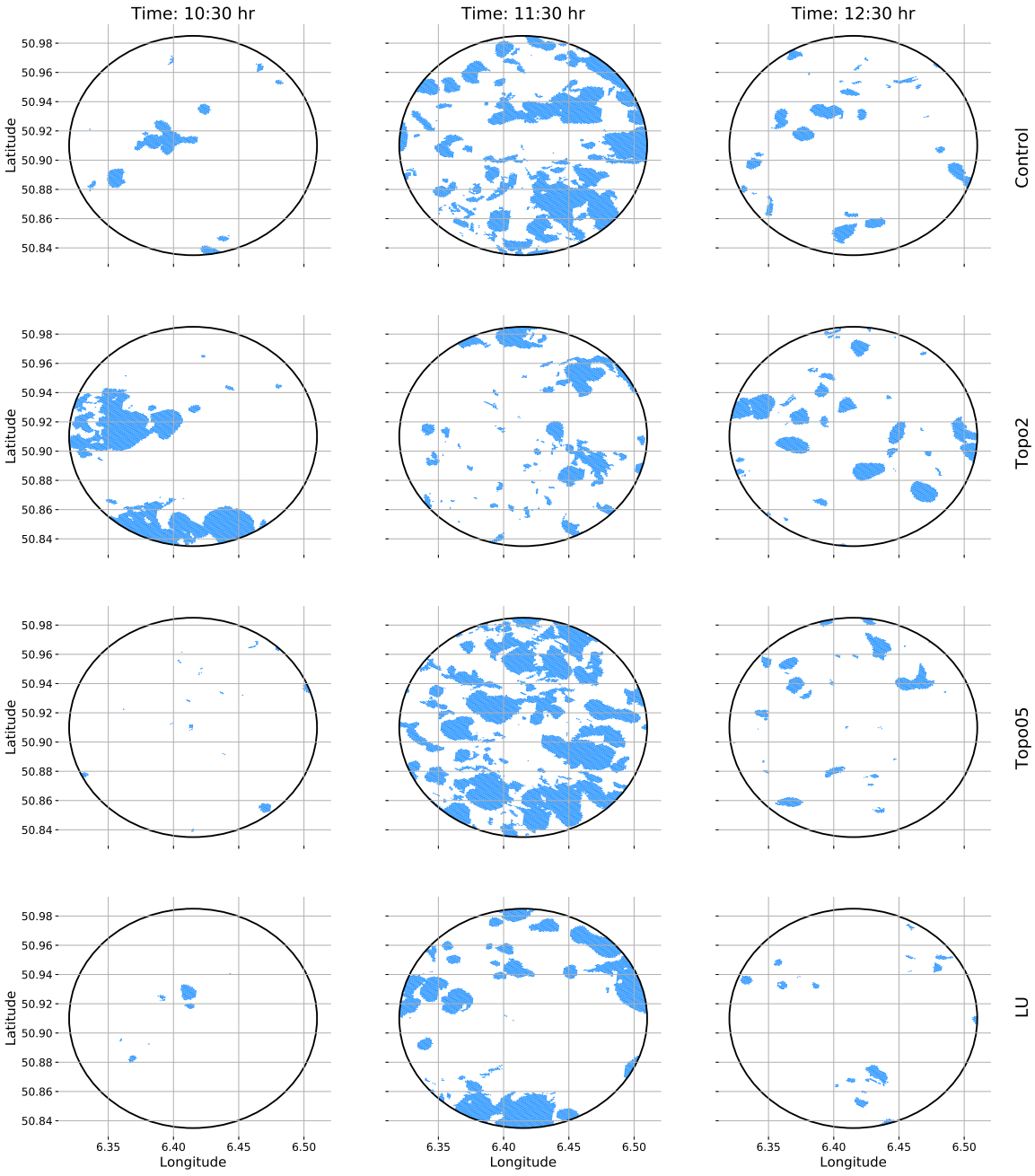

Figure 6.7: Cloud mask in blue for day A for three times during the day (columns) for the four different simulations (rows). The times are marked in Figure 6.5a as dashed lines.

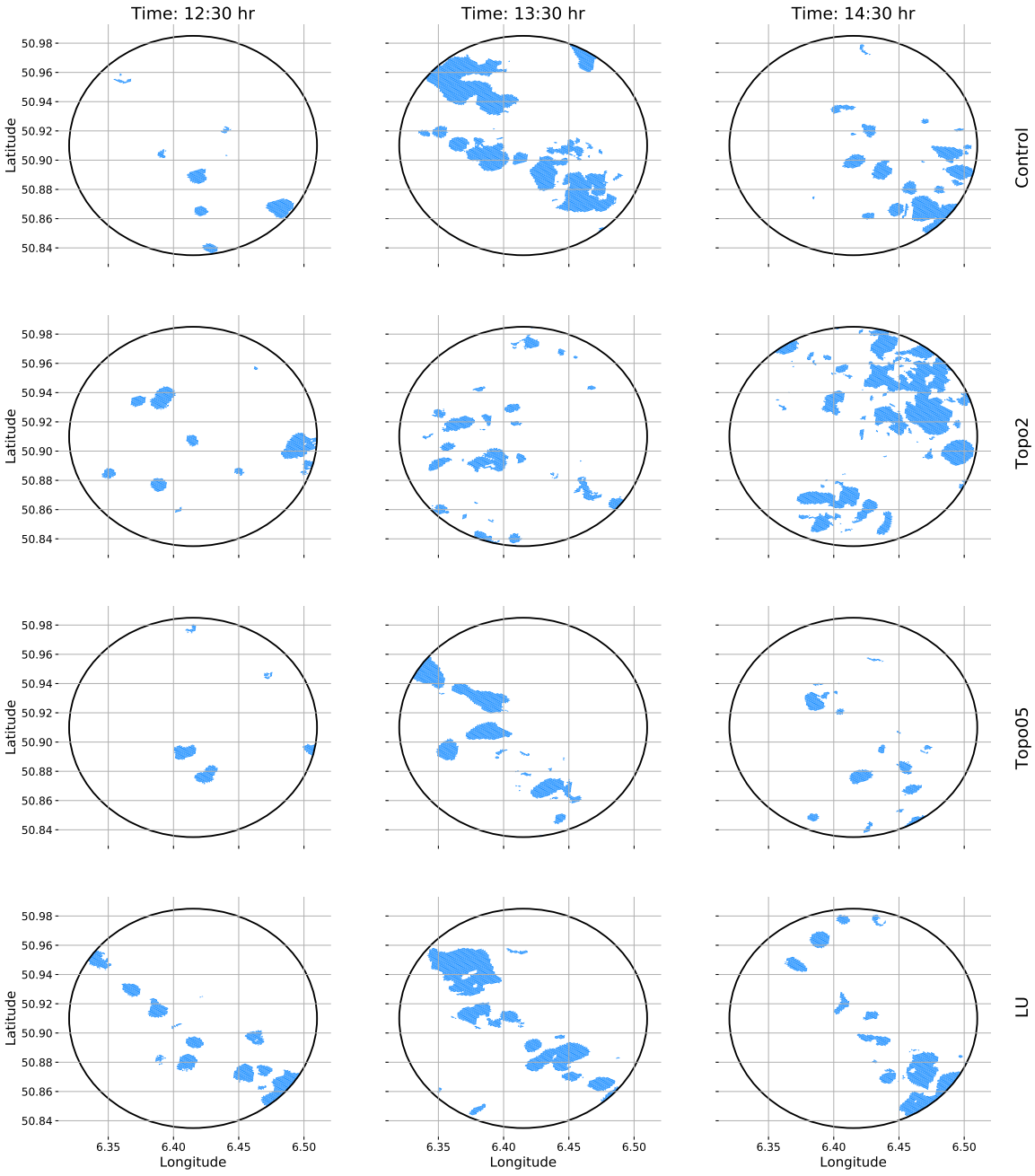

Figure 6.8: Cloud mask in blue for day B for three times during the day (columns) for the four different simulations (rows). The times are marked in Figure 6.5b as dashed lines.

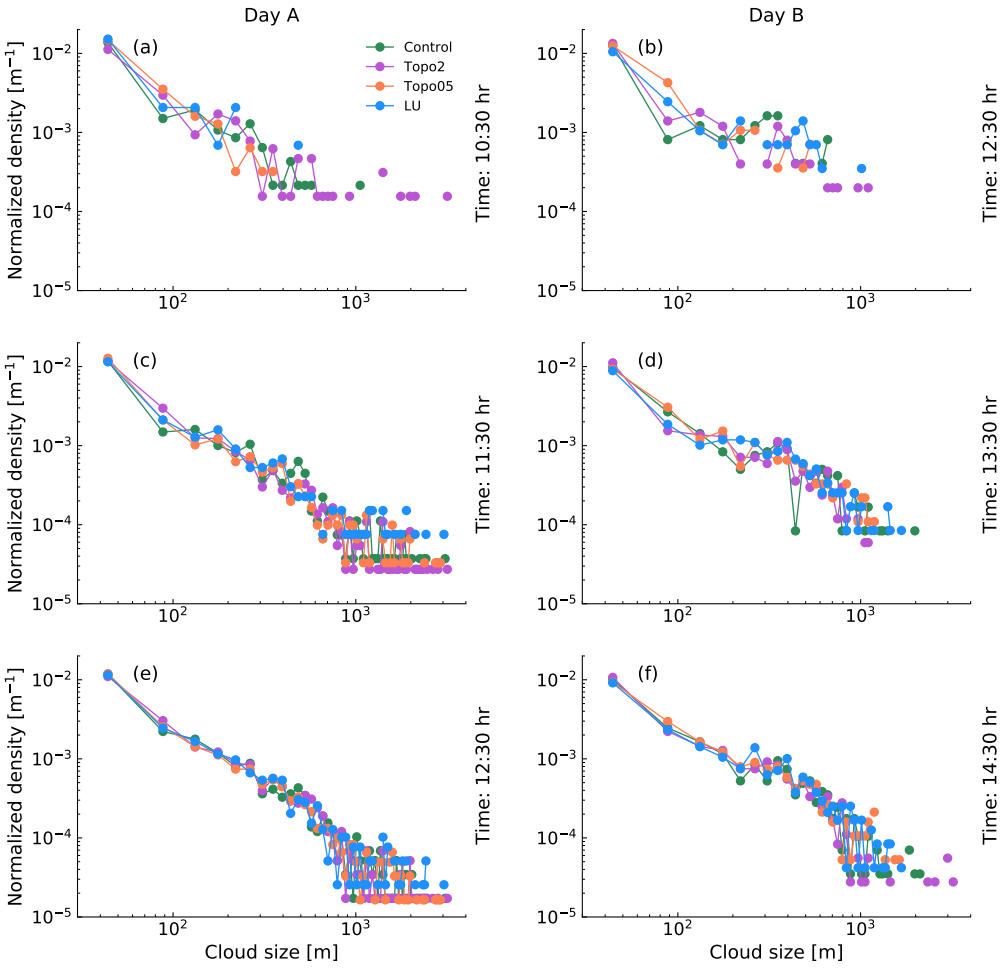

Figure 6.9: The normalized cloud size densities (CSD) for all simulations, the two days (columns), and the three timesteps that are also shown in Figures 6.7 and 6.8 (rows).

there are also some clouds present at the edges of the domain. An hour later, the number and size of clouds has decreased.

For a more quantitative way of describing the cloud field, the Cloud Size Distribution (CSD) is shown in Figure 6.9. The CSD is computed for all simulations and for the timesteps also shown in the snapshots of the cloud field (Fig. 6.7 and 6.8). For the computation of the CSD, all cloudy grid cells (containing liquid water) are clustered into clouds. The size of the cloud is defined as the radius of a circle with the same area as the cloud. This method is described in more detail in Chapter 4.

For day A we can see that over time the maximum cloud size increases for all simulations except Topo2. Topo2 has at 10:30 hr the largest clouds, and they stay present for all timesteps. The largest cloud present in Control is the smallest compared to the largest clouds in other simulations. This is not the case for day B. For Control, the largest cloud size is largest of all simulations at 13:30 hr and second largest at 14:30 hr. The simulations of day B show large differences in the CSD for the first timestep. However, for the last two timesteps the shape of the CSD is very similar for all simulations, over a large range of scales.

For both days it can be observed that Topo2 often shows the largest range of cloud sizes. This is also shown in the timeseries of the maximum cloud size (Fig. 6.10). Topo2 has an

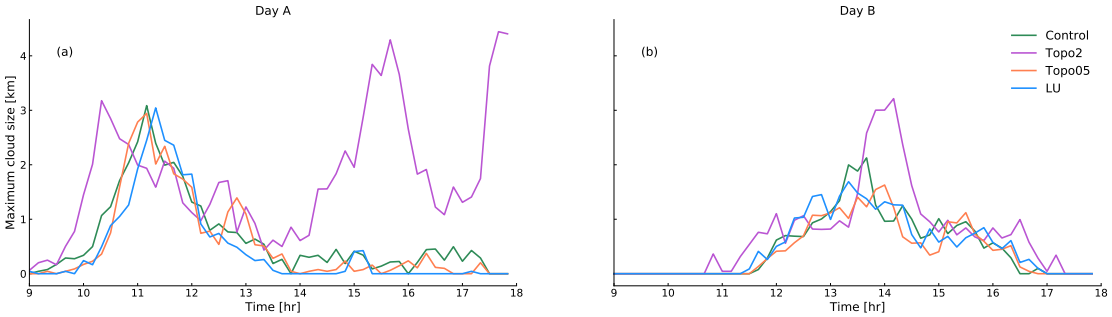

Figure 6.10: Timeseries of the maximum cloud size for all simulations.

early peak in maximum cloud size for day A and also has large clouds in the afternoon. The presence of clouds in the afternoon for Topo2 is analysed in the Appendix (section 6.6. For day B Topo2 also has the largest clouds, except for some hours in the morning and afternoon. The other 3 simulations do not show significant differences in maximum cloud size during the day.

For both days can be said that the shape of the CSD becomes more defined over time. The first part of a CSD is often defined as a power law (Neggers et al., 2003; Heus and Seifert, 2013), possibly extended to a power law-exponential function (Chapter 3). For our simulations the power law part of the CSD extends to larger scales as time progresses and the slope is very similar for all simulations. The slope becomes less steep over time. The better defined power law for the last timestep shown is interesting, since for this timestep there are not necessarily more clouds present. What we see is that with less clouds, and therefore more variability in the CSD, the relation is clearer defined and all simulations agree. These results suggest that some cloud related processes are similar for all simulations and that they play a role in determining the shape of the CSD as well. The differences between the simulations are thus not necessarily visible in a change of slope, but it is visible in the size of the largest cloud. This is in agreement with Rieck et al. (2014) who found that the effect of surface heterogeneity on the CSD is only visible in the scale break of the power law, and not in its slope.

6.3.3 Effect on organization

For the quantification of the organization of the clouds, I_{org} (Chapter 5 and Tompkins and Semie (2017)) is computed for all fields with more than five clouds (Fig. 6.11). I_{org} is an organization parameter based on nearest neighbor distances between clouds. The distances are compared to a theoretical random distribution (Poisson). Values higher than 0.5 indicate clustering of clouds. Also shown in Figure 6.11 are the number of clouds. The relation between I_{org} and the number of clouds is analysed to make sure I_{org} is not dominated by the number of clouds, as is the case for SCAI (5 and Tobin et al. (2012)), another organization parameter.

For day A, the time evolution of I_{org} (Fig. 6.11a) is similar for most simulations. When looking carefully, it could be observed that all simulations except Topo2 show a lower I_{org} for the time cloud fractions and the number of clouds is highest (Fig. 6.11c). This pattern is strongest for LU, with the lowest I_{org} value around 11:30 hr. At this time all clouds are present at the edges of the domain, leaving a hole in the middle (Fig. 6.7). For Topo05 this empty space in the middle of the domain is somewhat visible as well, which might cause the low I_{org} values between 11:00 and 11:30 hr. I_{org} is not significantly different in magnitude for the different simulations, even though the amount of clouds shows

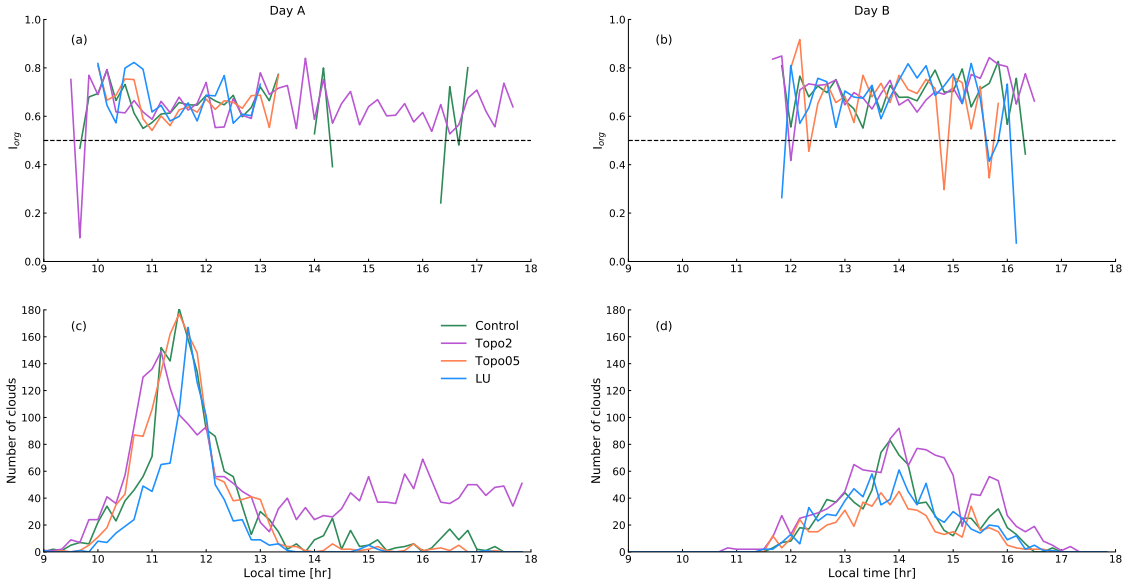

Figure 6.11: Timeseries of I_{org} for all four simulations for day A (a) and B (b). Timeseries of the number of clouds for all four simulations for day A (c) and B (d).

considerable differences. This means that regardless of the amount of clouds present, they always occur clustered, according to I_{org}. One could therefore say that the number of clouds do not directly influence the value of I_{org}, other than that a sufficient amount of them should be present.

Day B has less clouds than day A, but nevertheless a clear increase/decrease of number of clouds is happening over the course of a few hours (Fig. 6.11d). This trend is not visible in I_{org} (Fig. 6.11b). During the hours with clouds, I_{org} for all simulations for day B does not show a clear trend. A separation between the different simulations is visible only at 14:30 hr. LU shows here higher I_{org} values, it is around this time that the clouds form a cloud street, diagonally in the field (Fig. 6.8). This is apparently picked up by I_{org}. The simulation that does not show this pattern is Topo2. It has the lowest I_{org} values around this time, even though it exhibits highest number of clouds. Enhancing the topography thus does affect cloud formation, but hardly shows any difference in the degree of organization.

6.3.4 Cloud street formation

Even though I_{org} does not show clear differences between the simulations, we can still identify an interesting distribution of clouds from Figure 6.8. The formation of a cloud street, particularly visible for LU at 13:30 hr, hints at the presence of certain conditions that favour cloud formation in the middle of the domain. Analysing this situation might provide more insight in the underlying process of cloud organization and how to best quantify that. The cloud street is visible for both Control and LU. We will focus here on LU since that simulation has slightly higher I_{org} values in the afternoon.

As can be seen in Figure 6.8, the clouds are mainly located along a diagonal, north-west to south-east. This pattern is not visible for day A, an indication that more than land use change is necessary to create this organization. To provide a comparison between the LU simulations for the two different days, the surface sensible heat flux (SH_s) for 11:30 hr, day

A, is shown in Figure 6.12. Similarly, SH_s for 13:30 hr on day B can be seen in Figure 6.13. The difference in magnitude of SH_s is clear: in the middle of the domain day B features SH_s values of up to almost 500 W/m^2, whereas the maximum value for day A is around 200 W/m^2. Besides the magnitude also the amplitude of the differences in the domain is bigger for day B. The reason for this difference can be twofold: a difference in time of day and a difference in soil moisture availability. The snapshot of day B is two hours later than for day A, at this time the solar energy is higher, which could lead to higher heat fluxes. Different soil moisture contents result in a different partitioning of incoming solar energy into sensible and latent heat fluxes.

Added to the snapshots of SH_s are vectors for the surface horizontal wind. This shows two other differences between the simulations, coming from the large scale forcing: both the average wind speed and direction are different for the two days. Day A has low wind speeds, some circular cells of convergence and in the eastern part of the domain eastern winds, which are orthogonal to the heterogeneity pattern at the surface. Day B, on the other hand, has relatively strong winds coming from the north-west. In this simulation, clouds thus form parallel to the main wind direction, an observation also made by LeMone (1973).

Due to the cloud street in LU for day B, one would expect some kind of secondary circulation. To achieve this there should be convergence of horizontal wind towards the middle of the domain, orthogonal to the cloud street. This is hard to observe in Figure 6.13, but when looking at a vertical cross-section we can observe some patterns (Fig. 6.14). The cross-section is taken orthogonally to the cloud street and prevailing wind direction (following the black line in Fig. 6.13) and the quantities are averaged over one hour (13:30 to 14:30 hr). The vertical wind speed (Panel a) shows strong updrafts in the southern part of the cross-section, where there is also a high amount of moisture available (Panel b). The updrafts can be explained by a high SH_s (Panel c), but the high moisture contents do not seem to come from the surface, since the latent heat flux in the area of the clouds is low (Panel d). This low latent heat flux is understandable, considering the land use type in this area is bare ground. Because of the low moisture availability at the surface, most of the solar energy is released in the form of sensible heat. The humidity in this area is being transported from the region upstream, possibly outside the analysed domain.

A possibility that has to be studied further is the advection of the cloud street into the domain by the forcing at the boundary, especially since very similar structures are visible for Control at the same time. Topography seems to have an influence, for both diminished and enhanced topography the cloud street is less significant. This could indicate some influence of the surface conditions. However, further research is needed to assess the role of the forcing at the boundaries of the domain.

Judging from the averaged wind speed along the cross-section (Panel e), no full secondary circulation develops, since there is no convergence towards the middle of the cross-section. This might be because the background wind speed is too high (Lee et al., 2019). The analysis showed that the formation of a secondary circulation and a cloud street depends on a complex interplay of different processes. The realistic model set-up was able to capture these processes. Also, without idealized surface heterogeneity, a quasi-secondary circulation can be observed.

6.4 Alternative organization parameter

From the snapshots of cloud cover for the different days and simulations we have seen that the spatial distribution of clouds can vary significantly if surface conditions change.

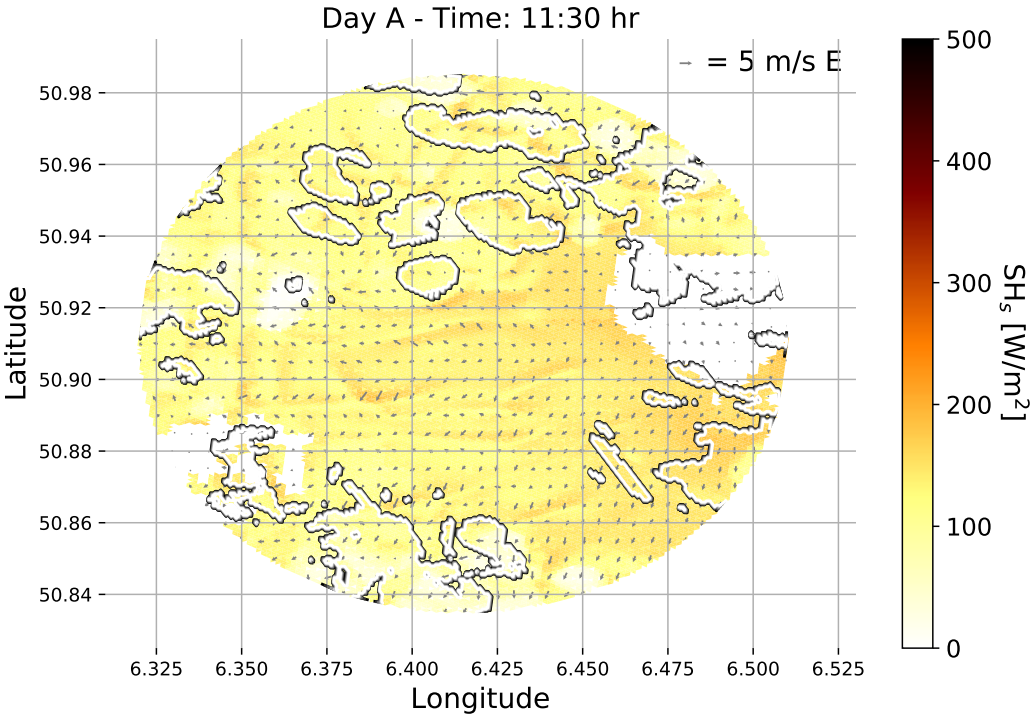

Figure 6.12: Surface sensible heat flux for 11:30 hr for day A (coloured). The black contours indicate cloud cover, the arrows show the horizontal wind speed and direction.

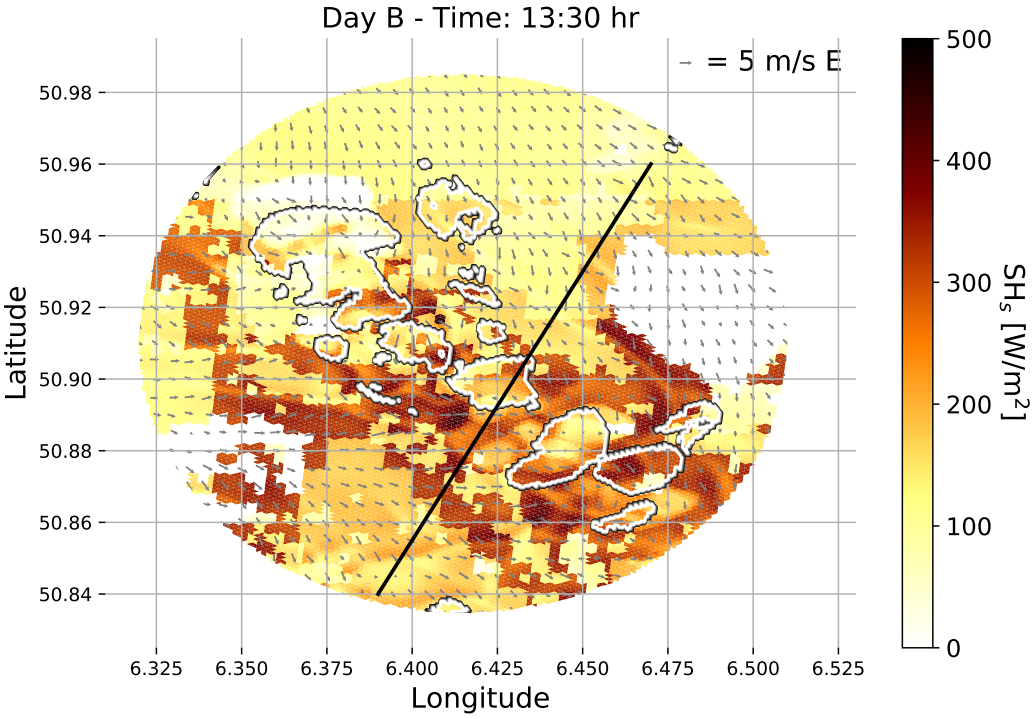

Figure 6.13: Same as Fig. 6.12, but for 13:30 hr day B. The black line shows the location of the vertical cross-section that is shown in Figure 6.14.

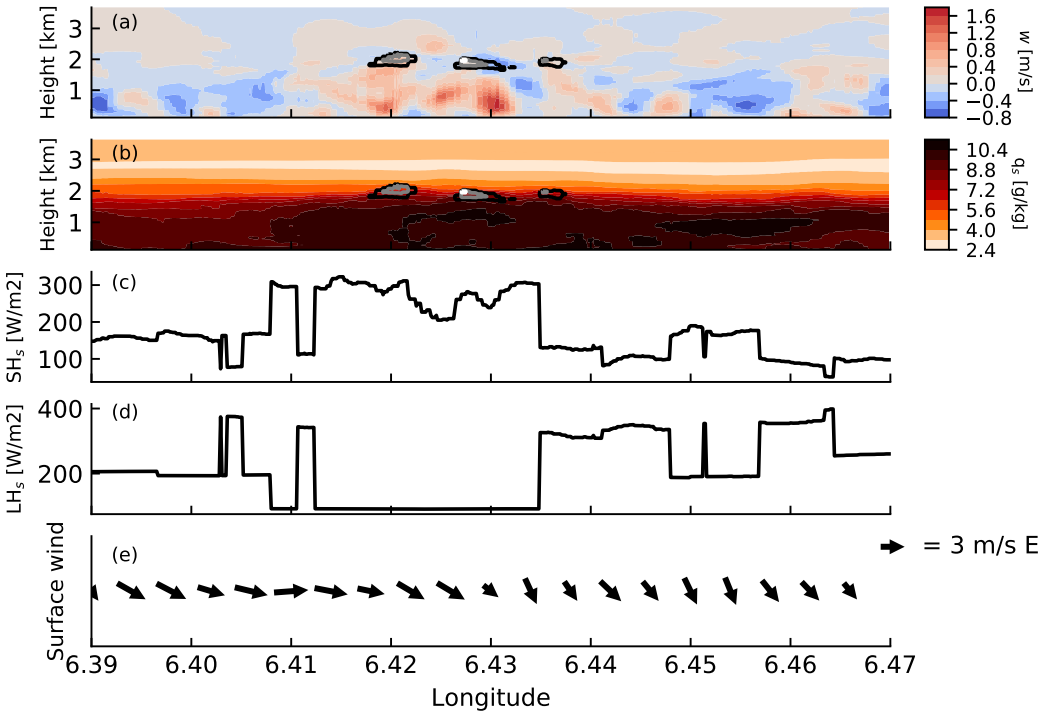

Figure 6.14: Contour for (a) vertical velocity (w) and (b) specific humidity (q_s). (c) The surface sensible (SH_s) and (d) latent heat flux (LH_s). (e) The wind direction and speed. All panels show the vertical cross-section along the line seen in Figure 6.13 and averaged over one hour (13:30-14:30 hr).

However, quantifying this behaviour is not straightforward. Using I_{org}, it is possible to distinguish between the different simulations, but the values themselves are not very indicative. A clear temporal evolution of I_{org} is not present; there are either no clouds, or I_{org} is very similar to all other timesteps and simulations. As soon as clouds are present, I_{org} indicates that they are clustered.

In the previous chapter (Chapter 5), we saw that I_{org} shows a trend of decreasing organization and a range of different values. That is not the case for the simulations analysed in this chapter. It might have several reasons. It could be that the cloud fields are just too similar to each other and that indeed their degree of organization is very similar. Another possible reason is the limited domain size and number of clouds in this study. The analysis showed that the number of clouds themselves do not seem to influence I_{org} directly, but it is possible that there is lower threshold for number of clouds and domain size that makes I_{org} informative.

I_{org} and the other organization parameters discussed previously are all object-based. However, there are also studies that suggest to look at organization in a different way. Bretherton et al. (2005) use the variance of total water as an indication of mesoscale aggregation. Lee et al. (2019) found that stronger organized systems exhibit a higher variance of vertical velocity. Taking into account the thermodynamic state of the boundary layer to assess organization could be an interesting approach, especially when looking at surface heterogeneity in relation to convective organization. In van Heerwaarden and de Arellano (2008) organization of convection happens when the amplitude of surface heterogeneity is strong enough. A very heterogeneous surface leads to large differences in atmospheric properties like temperature, moisture and vertical wind.

For an indication of how much information the variance of vertical velocity (w) could give,

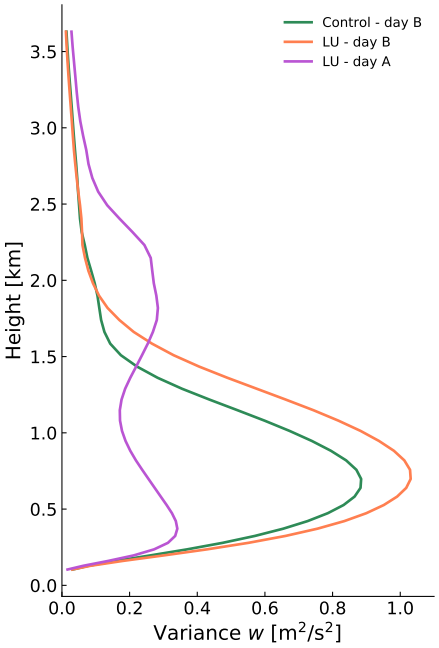

Figure 6.15: The vertical profile of the vertical wind speed variance. The profiles for day A (a) are averaged over 10:30-11:30 hr. The profiles for day B (b) are averaged over 13:30-14:30 hr.

it is shown as a vertical profile for the time of the formation of the cloud street for day B for Control and LU, and for the time of cloud presence for LU, day A. Comparing LU for day B to the other simulations, it can be seen that it has clearly more variance than LU for day A, but the differences to Control (which also shows a cloud street) are small. For day B, both profiles have a very similar shape, but the peak for LU at 750 m reaches slightly higher values.

Since there do seem to be some differences in the variance of w, the maximum of the vertical profile is determined for every simulation and every timestep. The timeseries of the maximum variance of w are shown in Figure 6.16. Day A shows for most simulations a diurnal evolution with the highest values between 12:00 and 13:00 hr. Topo05 shows a clear peak at 11:00 hr, although it has not the highest amount of clouds at this moment. LU shows over the whole day the lowest values, whereas in general Topo2 has the highest values. This could be connected to the fact that Topo2 features the largest clouds (Fig. 6.10). A higher variance in w might indicate stronger up- and downdrafts, giving the clouds the chance to grow bigger. The differences between the simulations are smaller for day B. All simulations show a peak at the time of the highest cloud cover (14:00-15:00 hr). Worth noting is the high variance for LU at 14:30 hr, this is the time where also I_{org} gives high values.

The variance of the vertical velocity shows a diurnal cycle that is not demonstrated for I_{org}. For day A, a distinction can be made between the different simulations, but the differences for day B are not significant. The variance seems to give some information on the convective flow in the domain, but in itself it might not be enough to indicate organization. A clear benchmark case is missing for our analysis, this complicates the evaluation of the variance of w. A combination of I_{org} and the variance of w could be considered for further research.

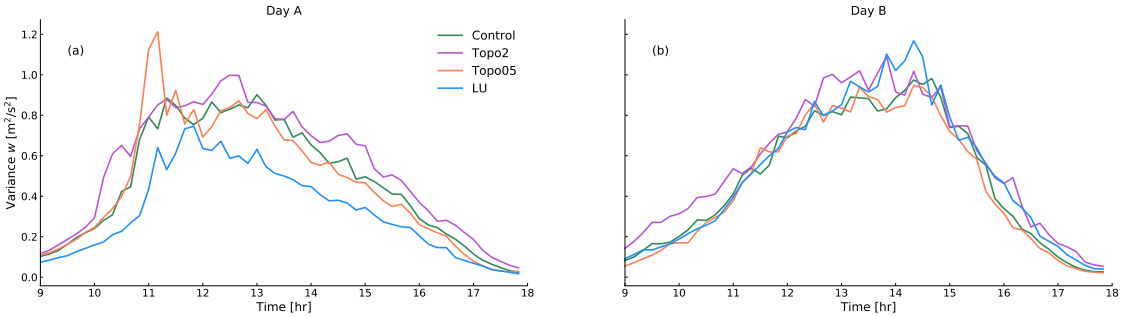

Figure 6.16: Timeseries of the maximum vertical velocity variance for all simulations for day A and B.

6.5 Discussion and conclusions

In this chapter we showed the impact of the surface conditions on the spatial distribution of shallow cumulus clouds. Changing the topography and land use type distribution caused differences in when and where clouds formed. This was analysed by using CSDs and organization parameter I_{org}.

Changing the surface conditions had little effect on the shape of the CSD. The main changes could be observed for the range the CSD covered, Topo2 often showed the largest range and Topo05 the smallest. The absence of strong topography and a patchy distribution of land use types for Topo05 diminished the chance of the formation of secondary circulations or strong updrafts. This inhibits the formation of larger clouds.

Topo05 also shows generally a low number of clouds, although this does not seem to affect the degree of organization. Unlike SCAI, for small domains I_{org} is not dominated by the number of clouds. Topo2 has more clouds in general for both simulated days as well as larger maximum cloud sizes. Yet it does not show higher I_{org} values. Interestingly, at the times where the number of clouds starts to decline, the CSD seems to be better defined among the simulations. The change in slope over time is similar for the simulations, which suggests a similar development of the cloud population.

For one of the simulated days a cloud street was formed. This can be explained by higher surface heat fluxes in the area of the cloud street and the transport of moisture towards this area. In a cross-section averaged over one hour we could observe the development of a quasi-secondary circulation. To increase the sample size, it would be good to also average spatially, i.e. using more than just one slice. To more accurately determine the role of the surface in the formation of the secondary circulation additional simulations have to performed. These simulations should have homogeneous surface conditions or a perturbed initialization. Also the forcing data at the boundary of the domain should be analysed in order to establish the origin of the secondary circulation.

Because of the influence of secondary circulations on cloud organization, it sounds reasonable to incorporate variables that reflect this process in a metric to quantify organization. As a simple experiment timeseries of maximum w variances were computed. These timeseries give a different signal than I_{org}. There is a clear temporal evolution and some differences between the simulations could be distinguished. Just as for the CSD, the differences between the simulations are smaller for day B than for day A. The fact that the variance of w shows differences between simulations could indicate that there might be more differences than shown by I_{org}. However, for an evaluation of the variance of w as an indicator of organization it would be useful to start with some idealized simulations. A

clear benchmark case is missing in this study, it would be helpful to have a case where the degree of organization is known.

In this chapter we showed that also for realistic simulations it is possible to detect the impact of surface heterogeneity and the development of secondary circulations. A change in topography or land use can alter cloud formation. Although we can see from snapshots that the spatial distribution of clouds differs among the simulations, quantifying this behaviour still proves ambiguous. This is also complicated by the fact that the reasons for spatial distributions might be present outside the domain that is observed (see section 6.6). For further development of any object-based approaches satellite observations might prove useful. Having high-resolution imagery of patterns that occur in nature can help in developing a useful tool. When using variables that describe the state of the boundary layer to quantify organization, first disentangling the different processes involved might be necessary. That way the role of the surface might become more clear and can be accounted for properly.

6.6 Appendix: cloud cover for enhanced topography

The analysed results indicate a change in cloud cover for the Topo2 simulations. Large maximum cloud sizes (Fig. 6.10), many clouds (Fig. 6.11) and relatively high values for the variance of vertical velocity (Fig. 6.16) all hint at some stronger convection for Topo2. For a more detailed analysis of the role that topography plays in this we will here focus on Control and Topo2 of day A. More specifically on the afternoon of the Topo2 simulation. As can be seen from the snapshots (Fig. 6.7) and timeseries (Fig. 6.11), Topo2 has less clouds than Control between 11:00 and 12:00 hr. However, in the afternoon Topo2 is the only simulation for day A that is still showing clouds.

The peak in cloud cover in the afternoon for Topo2 is between 15:00 and 16:00 hr (Fig. 6.17a). Looking at the domain averaged heat fluxes there is no significant difference between Control and Topo2 (Fig. 6.17b). Where the cloud formation in the morning hours can be related back to the variance of the heat fluxes, this is not the case for the afternoon (Fig. 6.17c). Thermodynamics therefore do not seem to explain the large difference in cloud cover. The horizontal wind speed, averaged over the domain and time (15:00-16:00 hr, Fig. 6.19) shows some small differences between Control and Topo2. Over the whole boundary layer Control has slightly higher wind speeds than Topo2. The difference is small though, it seems unlikely that this difference in wind speed explains the large difference in cloud cover.

Since the domain averaged values give no clear indication of what could cause the difference in cloud cover, the spatial distribution of these values might give some more insight. Most clouds for Topo2 at 15:30 hr are located in the southern part of the domain (Fig. 6.18). To compare the spatial distribution of the heat fluxes, the values for Control are subtracted from Topo2. For both the sensible heat flux (Fig. 6.20) and the latent heat flux (Fig. 6.21) there are negative values present in the southern part of the domain. This means that in the area where clouds are present the fluxes of Topo2 are lower than for Control. In this area, the vertical wind is predominantly downwards as opposed to the upward vertical wind in the more northern area (not shown). The difference in horizontal wind speed and direction is represented by the wind vectors in Figures 6.20 and 6.21, but they do not show

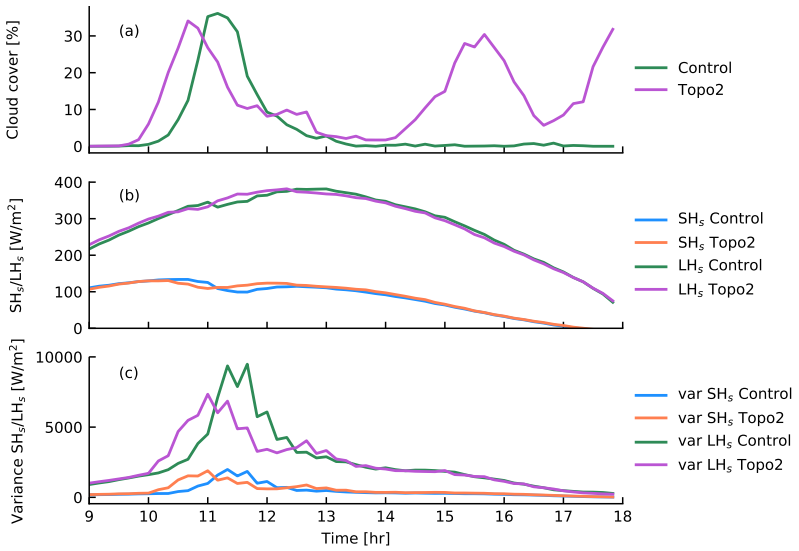

Figure 6.17: Timeseries for Control and Topo2 for day A of cloud cover (a), surface sensible and latent heat flux (b) and the variance of these heat fluxes over the domain (c).

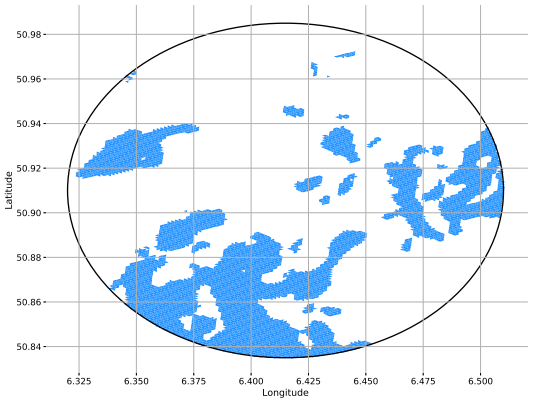

Figure 6.18: The cloud cover in blue at 15:30 hr for Topo2 of day A.

a clear pattern that could explain the difference in cloud formation. When comparing the southern area with clouds to the northern area without, the vertical distribution of temperature and humidity does not differ (not shown).

In previous studies on the cloud distribution over heterogeneous surfaces it is often found that clouds form over the warm and dry patches because of the strong upward motion that is often associated to them. The necessary moisture is transported from the neighboring patch (Avissar and Schmidt, 1998; Kang and Bryan, 2011; Lee et al., 2019). However, in our case both the sensible and the latent heat flux are lower than for the simulations without clouds and less energy is transported upwards from the surface. Therefore the spatial distribution does not explain the location of the clouds. While it is unclear where the clouds come from, the lower heat fluxes can be explained by cloud shading. The presence of clouds blocks incoming solar radiation thereby lowering the energy available for the surface fluxes. This is also observed by e.g. Rieck et al. (2014).

The analysis of this case shows that although the cloud field between two very similar simulations might be very different, a simple explanation is not always given. In this particular case the reason might well be located outside the domain that we analyse. The complete simulation set-up consists of four nested domains, all of them have enhanced topography for Topo2. The clouds in Topo2 might originate from conditions in one of the outer domains. Capturing this behaviour in an organization parameter that is based on local conditions is therefore not straightforward.

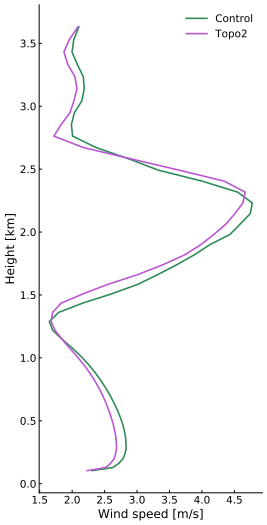

Figure 6.19: The vertical profile of horizontal wind averaged over the domain and over 15:00-16:00 hr for Control and Topo2 of day A.

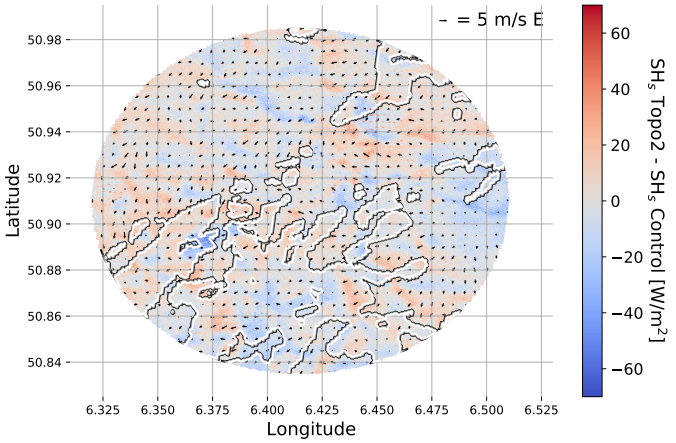

Figure 6.20: The surface sensible heat flux for Topo2 minus the surface sensible heat flux for Control for day A. The arrows indicate the difference in wind speed and direction between Topo2 and Control (Topo2-Control).

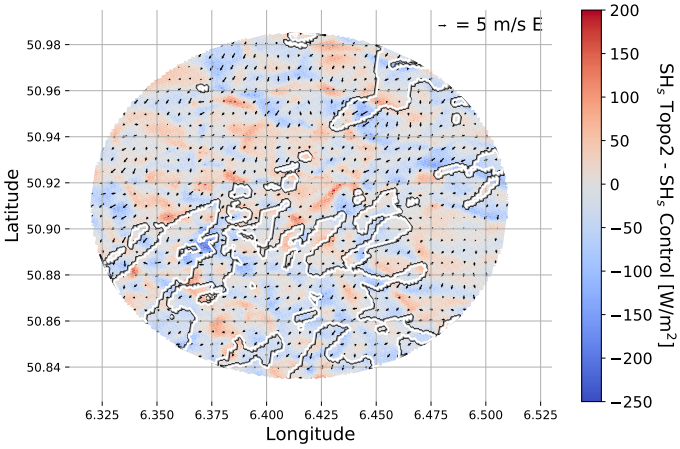

Figure 6.21: Same as Fig. 6.20 but now for the latent heat flux. Note the difference in range of values.

Chapter 7

Conclusions and Outlook

The goal of this thesis is to describe shallow cumulus cloud populations in terms of size and spacing and to analyse how they interact with the surface. With the achieved results of this work we are now able to answer the research question posed in the motivation. The increased understanding of cloud populations gives directions for further research as well.

7.1 Answering the research question

In this thesis shallow cumulus cloud populations are simulated in varying set-ups. Their spatial patterns are characterized by applying several different tools. Making use of an ensemble of Large Eddy Simulations, their typical size distribution is studied. Their typical spacing is analysed with data from large-domain simulations over the subtropical ocean. All of this to finally be able to assess the impact of surface conditions on the spatial distribution of shallow cumulus clouds.

To describe a shallow cumulus cloud population in terms of size, it is found that the cloud size distribution is well captured by a power-law exponential function. This function describes the distribution throughout the full range of scales. Thanks to the library of cases and the large parameter space because of that, we could show that the largest cloud size present in the domain correlates with the total cloud cover. Cloud size also plays a role in cloud spacing. Large clouds have a large spacing with respect to their closest neighbor, as opposed to small clouds which have a small spacing. Spatially, the more numerous small clouds surround the larger ones.

For actual use of information on the spatial distributions of clouds, quantification of the patterns that can be observed is of importance. To this end, several organization parameters are compared. All parameters are able to distinguish between randomly distributed and organized cloud fields. The parameter giving most information (I_{org}) is then used to assess the influence of surface heterogeneity on spatial patterns in cloud populations, as we set out to show from the beginning. Namely, the research question at the basis of this thesis is:

How does surface heterogeneity influence the spatial pattern of shallow cumulus clouds?

To answer this question, a sensitivity study is performed which included a modification of the amplitude of topography and land use type distribution. The study shows an increase of cloud cover and larger clouds for enhanced topography. For the simulations with a

modified distribution of land use types, the cloud pattern is altered by the formation of a cloud street. The formation of a secondary circulation that could explain the cloud street is influenced by the surface conditions, both the land use distribution and the topography. I_{org} classifies the cloud fields for all simulations and at all timesteps as clustered. Based on I_{org} no clear distinction between the simulations can be made, even though some differences in the cloud spatial distribution are reflected in I_{org}. From the cloud size distributions could be observed that simulations with enhanced topography showed generally larger clouds.

7.2 Further research possibilities

The ultimate goal considering convective organization would be to include it in a parametrization scheme. Organization can be included directly, by using an organization parameter, or indirectly. To include organization indirectly, the effects of organization on the development of cloud populations should be accounted for in the parametrization scheme. An analysis of the combined results of this thesis might give some hints for further research on this topic.

7.2.1 Differences over land and ocean

For the comparison of different organization parameters, simulations were used that were performed over the subtropical ocean, meaning that the surface was relatively homogeneous. For this data, a clear transition from organized to non-organized convection could be observed. This transition was reflected in the time evolution of I_{org}. The simulations performed over land to assess the influence of surface heterogeneity were very different than the one over the subtropical ocean, which could explain the different behavior of I_{org}. The analysed domain was smaller and the resolution higher, but maybe most important, the simulations were performed over land. As explained in Chapter 2, over land the driver for atmospheric convection is strongly coupled to the time of day. For these simulations I_{org} turned out to be less informative.

There could be several reasons for the difference in performance of I_{org}, as both spatial and temporal scales could play a role. Simulations done on a domain that is too small might have not enough space for organization to develop as much as needed for I_{org} to detect it. Increasing the domain size could help to gain more insight on this. Since organization over land is strongly coupled to the diurnal cycle there is limited time for the organization to develop. Simulations with prolonged day time hours could be useful to study the time scales needed for organization to establish.

An important extra simulation to perform is over land, but with homogeneous surface conditions. Results from this simulation are important in two regards. First, it would give the possibility to assess the influence of the diurnal cycle of heat fluxes when compared to the simulation performed over the ocean, which has a much less pronounced diurnal cycle. Second, not only the effect of land surface, but also of land surface heterogeneities can be more explicitly studied.

7.2.2 How to study organization

An accurate description of organization would aid the development of parametrization schemes, especially scale-aware schemes. An organization parameter that is applicable over the whole range of scales of convection would be necessary for that. In general, one could state that an organization parameter has to fulfil two prerequisites; the variable

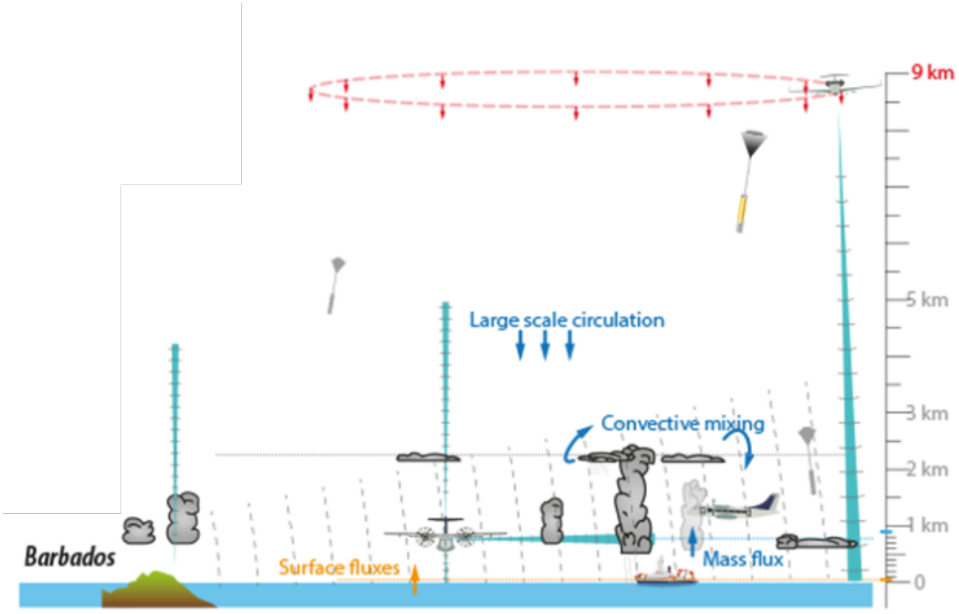

Figure 7.1: Overview of the measurement set-up for the EUREC^{4}A campaign.
Source: http://eurec4a.eu/

on which it is based should be observable and the parameter should be applicable for parametrization schemes. In this study we have seen that the object-based approach fulfils both of these two requirements. Besides object-based approaches one could also apply a more physics-based approach. Here we will discuss the two options.

An important advantage of using object-based approaches is that they can be directly applied to satellite observations. Satellite measurements have increased in horizontal resolutions and they are now in the range of resolving the scales of shallow cumulus clouds. A one-to-one comparison is therefore possible. Quantifying organization based on an image also opens ways to incorporate methods and techniques from different fields of science, e.g. machine learning and image processing techniques.

From a simple conceptual model, as presented in Chapter 4, we have learned that the relation between cloud size and cloud spacing could be explained by distinguishing small and large clouds. We were able to reproduce the results from the simulations by assuming that small clouds are randomly distributed in the domain whereas large clouds are more organized. Different spacing behavior for different cloud sizes could be important for scale-aware parametrization schemes as well. One could therefore also think of a scale-aware organization parameter. The degree of organization as a function of cloud size could then be directly reflected in the cloud size distribution (Neggers et al., 2019).

As mentioned in Chapter 5, normalizing the organization parameters with values for a random field (either 'self-made' or by using the Poisson distribution) is necessary to be able to interpret the results. For the computation of the random fields it is assumed that a cloud can be represented by a single point, instead of it being an object with a finite size. For the current study a small sensitivity test did not indicate an influence of this assumption on the results. However, it is important to have a clear benchmark case of randomness to be able to categorize a field as clustered or not. How to compute a truly random field while taking cloud sizes into account should therefore be considered a topic of further research.

An example of a physics-based approach is the exploration of the variance of vertical

velocity in Chapter 6. Is is known that organization influences humidity and precipitation. However, a direct relation between humidity and degree of organization is not established yet (Wing, 2019). Considering the direct impact of organization on its environment it is a relation that should be given attention. An advantage of an approach based on thermodynamical variables is the possibility to directly couple organization to prognostic variables in the model. Also would it be possible to relate these variables to surface conditions, since we have seen in Chapter 6 that surface heterogeneity plays a role in convective organization as well. The ongoing studies on cold pool dynamics could be of interest for this matter, as it relates humidity and precipitation to organization. Humidity might also be of importance for the lateral growth of clouds, it is hypothesized in Chapter 3 that humidity levels influence the maximum cloud size. Maximum cloud sizes change for different surface conditions as well (Chapter 6) and could therefore be an indication for change in spatial patterns.

7.2.3 The role of observations

So far the organization parameters have mainly been studied by using model data. Comparing the results of this thesis to observational data is a topic of further research. Both local measurements (e.g. from JOYCE) as well as regional scale observations from satellite data could help in understanding the dynamics of shallow cumulus populations. Also measurement campaigns can play an important role in developing our understanding of clouds and convection, especially when they are designed in a way that makes a comparison with model data possible. An example of a campaign like that is EUREC4A, planned to take place early 2020. During this campaign, both ships and air planes will measure over the subtropical ocean, near Barbados. A sketch of the set-up is given in Figure 7.1. The extensive measurements are focused on the interaction between clouds and their large-scale environment and on how this affects convective organization. A possible way to include the measurements in modelling studies is to compare the simulations to the observational results, or use dropsondes to nudge the simulations towards them (e.g. Reilly et al. (2019)).

Acknowledgements

During my adventure in Cologne I am supported by many people and I am truly thankful for that. First of all I would like to thank my supervisor Roel. Your never ceasing enthusiasm for the atmospheric sciences really is an inspiration and our discussions always helped me forward. I appreciate the freedom you gave me to try out new things and that you gave me opportunities to develop myself as a scientist. Bedankt!

I also would like to thank Susanne and Vera as members of my doctoral committee, your interest and investment in my work has been very motivational. Of course this work would not have been possible without the TR32 and the IRTG. Thanks to all the project members who helped me by giving structure and providing funding for travel and networking.

When working for years on a specific topic, with all the ups and downs that it entails, a pleasant daily work environment is absolutely a blessing. I would especially like to thank all the members of InScAPE: office mates Sara and Philipp, Vera, Jan, Stephanie and Maren. All of you helped me in one way or another, be it through proofreading, translating or being excellent travel companions. Go InScAPE!

Ik zou echter helemaal niet in staat zijn om dit te schrijven zonder mijn ouders. Paps en mams, jullie hebben mij altijd alle vrijheid gegeven om te doen wat mij het beste lijkt. Tegelijkertijd tonen jullie ook altijd interesse en weet ik mij zeker van jullie onvoorwaardelijke steun. Ik ben een gezegend mens met ouders als jullie. Natuurlijk kan ook de rest van mijn familie hier niet ontbreken, in het bijzonder mijn broers en zussen. Mijn bezoekjes aan Nederland waren dankzij jullie gezellige chaos altijd een welkome afleiding van werk en leven in Keulen!

Also to all my friends, both close by and far away, old friends and new friends, thanks for your support! I am very grateful to have you in my life and that you make my life fun! I saved the most special friend of all for the end. José Luis, to use your real name for once, I will always be thankful for you by my side during this journey. Our inspiring discussions about basically everything broadened my horizon, not only scientifically. Thanks for always be willing to answer my questions, I believe the blackboard in our kitchen provides enough evidence for this...

Bibliography

Arakawa, A. and Schubert, W. H. (1974). Interaction of a cumulus cloud ensemble with the large-scale environment, part I. *J. Atmos. Sci.*, 31(3):674–701.

Arakawa, A. and Wu, C.-M. (2013). A unified representation of deep moist convection in numerical modeling of the atmosphere. Part I. *J. Atmos. Sci.*, 70(7):1977–1992.

Avissar, R. and Schmidt, T. (1998). An Evaluation of the Scale at Which Ground-Surface Heat Flux Patchiness Affects the Convective Boundary Layer Using Large-Eddy Simulations. *J. Atmos. Sci.*, 55:2666–2689.

Banta, R. M. (1984). Daytime boundary-layer evolution over mountainous terrain. part 1: Observations of the dry circulations. *Mon. Wea. Rev.*, 112(2):340–356.

Barker, H. W. and Räisänen, P. (2005). Radiative sensitivities for cloud structural properties that are unresolved by conventional GCMs. *Q. J. Roy. Meteorol. Soc.*, 131(612):3103–3122.

Benner, T. C. and Curry, J. A. (1998). Characteristics of small tropical cumulus clouds and their impact on the environment. *Geophys. Res. Lett.*, 103:28753–28767.

Betts, A. K. (2009). Land-surface-atmosphere coupling in observations and models. *J. Adv. Model Earth Syst.*, 1(3):n/a–n/a.

Bontemps, S., Herold, M., Kooistra, L., et al. (2012). Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences*, 9(6):2145–2157.

Bony, S. and Dufresne, J.-L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.*, 32(20):L20806.

Borque, P., Kollias, P., and Giangrande, S. (2014). First observations of tracking clouds using scanning ARM cloud radars. *J. Appl. Meteor. Climatol.*, 53(12):2732–2746.

Brast, M., Schemann, V., and Neggers, R. A. J. (2018). Investigating the scale adaptivity of a size-filtered mass flux parameterization in the gray zone of shallow cumulus convection. *J. Atmos. Sci.*, 75(4):1195–1214.

Bretherton, C. S. and Blossey, P. N. (2017). Understanding mesoscale aggregation of shallow cumulus convection using large-eddy simulation. *J. Adv. Model Earth Syst.*, 9(8):2798–2821.

Bretherton, C. S., Blossey, P. N., and Khairoutdinov, M. (2005). An energy-balance analysis of deep convective self-aggregation above uniform SST. *J. Atmos. Sci.*, 62(12):4273–4292.

Brient, F., Schneider, T., Tan, Z., et al. (2015). Shallowness of tropical low clouds as a predictor of climate models' response to warming. *Clim. Dyn.*, 47(1):433–449.

Brown, A. R., Cederwall, R. T., Chlond, et al. (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Q. J. Roy. Meteorol. Soc.*, 128(582):1075–1093.

Brune, S., Kapp, F., and Friederichs, P. (2018). A wavelet-based analysis of convective organization in ICON large-eddy simulations. *Q. J. Roy. Meteorol. Soc.*, 144(717):2812–2829.

Cahalan, R. F., Ridgway, W., Wiscombe, W. J., et al. (1994). The albedo of fractal stratocumulus clouds. *J. Atmos. Sci.*, 51(16):2434–2455.

Cioni, G. and Hohenegger, C. (2017). Effect of soil moisture on diurnal convection and precipitation in large-eddy simulations. *J. of Hydrometeorology*, 18(7):1885–1903.

Cohen, B. G. and Craig, G. C. (2006). Fluctuations in an equilibrium convective ensemble. part II: Numerical experiments. *J. Atmos. Sci.*, 63(8):2005–2015.

Coppin, D. and Bony, S. (2018). On the interplay between convective aggregation, surface temperature gradients, and climate sensitivity. *J. Adv. Model Earth Syst.*, 10(12):3123–3138.

Cronin, T. W. and Wing, A. A. (2017). Clouds, circulation, and climate sensitivity in a radiative-convective equilibrium channel model. *J. Adv. Model Earth Syst.*, 9(8):2883–2905.

Dal Gesso, S., van der Dussen, J. J., Siebesma, A. P., et al. (2015). A single-column model intercomparison on the stratocumulus representation in present-day and future climate. *J. Adv. Model Earth Syst.*, 7(2):617–647.

Dawe, J. and Austin, P. (2013). Direct entrainment and detrainment rate distributions of individual shallow cumulus clouds in an LES. *Atmos. Chem. Phys*, 13:7795–7811.

Dawe, J. T. and Austin, P. H. (2012). Statistical analysis of an LES shallow cumulus cloud ensemble using a cloud tracking algorithm. *Atmos. Chem. Phys*, 12:1101–1119.

de Roode, S. R., Duynkerke, P. G., and Jonker, H. J. J. (2004). Large-eddy simulation: How large is large enough? *J. Atmos. Sci.*, 61(4):403–421.

Ding, B., Li, C., Zhang, M., Lu, G., and Ji, F. (2014). Numerical analysis of percolation cluster size distribution in two-dimensional and three-dimensional lattices. *The European Physical Journal B*, 87(8):179.

Dipankar, A., Stevens, B., Heinze, R., et al. (2015). Large eddy simulation using the general circulation model ICON. *J. Adv. Model Earth Syst.*, 7(3):963–986.

Dorrestijn, J., Crommelin, D. T., Siebesma, A. P., and Jonker, H. J. J. (2013). Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theor. Comput. Fluid Dyn.*, 27(1):133–148.

Dufresne, J.-L. and Bony, S. (2008). An assessment of the primary sources of spread of global warming estimates from coupled atmosphere–ocean models. *J. Climate*, 21(19):5135–5144.

Eastman, R., Warren, S. G., and Hahn, C. J. (2011). Variations in cloud cover and cloud types over the ocean from surface observations, 1954–2008. *J. of Climate*, 24(22):5914–5934.

ECMWF (2017). *Part IV: Physical processes*. Number 4 in IFS Documentation. ECMWF.

Feingold, G., Balsells, J., Glassmeier, F., et al. (2017). Analysis of albedo versus cloud fraction relationships in liquid water clouds using heuristic models and large eddy simulation. *J. Geophys. Res.*, 122(13):2017JD026467.

Garcia-Carreras, L. and Parker, D. (2011). What is the Mechanism for the Modification of Convective Cloud Distributions by Land Surface-Induced Flows? *J. Atmos. Sci.*, 68:619–634.

Gesso, S. D. and Neggers, R. A. J. (2018). Can we use single-column models for understanding the boundary layer cloud-climate feedback? *J. Adv. Model Earth Syst.*

Glassmeier, F. and Feingold, G. (2017). Network approach to patterns in stratocumulus clouds. *Proceedings of the National Academy of Sciences*, 114(40):10578–10583.

Gustafson, W. I., Vogelmann, A. M., Cheng, X., et al. (2017). Description of the LASSO alpha 1 release. Technical report, University of California, Los Angeles; Jet Propulsion Laboratory.

Haerter, J. O., Berg, P., and Moseley, C. (2017). Precipitation onset as the temporal reference in convective self-organization. *Geophys. Res. Lett.*, 44(12):6450–6459.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York.

Heinze, R., Dipankar, A., Henken, C. C., et al. (2017). Large-eddy simulations over germany using ICON: a comprehensive evaluation. *Q. J. Roy. Meteorol. Soc.*, 143(702):69–100.

Heinze, R., Moseley, C., Böske, L. N., et al. (2016). Evaluation of large-eddy simulations forced with mesoscale model output for a multi-week period during a measurement campaign. *Atmospheric Chemistry and Physics Discussions*, pages 1–37.

Heus, T. and Seifert, A. (2013). Automated tracking of shallow cumulus clouds in large domain, long duration large eddy simulations. *Geosci. Model Dev.*, 6(4):1261–1273.

Heus, T., van Heerwaarden, C. C., Jonker, H. J. J., et al. (2010). Formulation of the dutch atmospheric large-eddy simulation (DALES) and overview of its applications. *Geosci. Model Dev.*, 3(2):415–444.

Hill, G. E. (1974). Factors controlling the size and spacing of cumulus clouds as revealed by numerical experiments. *J. Atmos. Sci.*, 31(3):646–673.

Hiraoka, Y., Nakamura, T., Hirata, A., et al. (2016). Hierarchical structures of amorphous solids characterized by persistent homology. *Proceedings of the National Academy of Sciences*, 113(26):7035–7040.

Hohenegger, C. and Bretherton, C. S. (2011). Simulating deep convection with a shallow convection scheme. *Atmospheric Chemistry and Physics*, 11(20):10389–10406.

Holton, J. R. and Hakim, G. J. (2012). *An Introduction to Dynamic Meteorology*. Academic Press.

Homar, V., Gayà, M., Romero, R., Ramis, C., and Alonso, S. (2003). Tornadoes over complex terrain: an analysis of the 28th august 1999 tornadic event in eastern spain. *Atmospheric Research*, 67-68:301–317.

Honnert, R. (2016). Representation of the grey zone of turbulence in the atmospheric boundary layer. *Advances in Science and Research*, 13:63–67.

Horn, G., Ouwersloot, H., de Arellano, J. V.-G., and Sikma, M. (2015). Cloud Shading Effects on Characteristic Boundary-Layer Length Scales. *Boundary-Layer Meteorol.*

Huang, H. and Margulis, S. (2013). Impact of Soil Moisture Heterogeneity Length Scale and Gradients on Daytime Coupled Land-Cloudy Boundary Layer Interactions. *Hydrol. process.*, 27:1988–2003.

Jakub, F. and Mayer, B. (2017). The role of 1-d and 3-d radiative heating in the organization of shallow cumulus convection and the formation of cloud streets. *Atmospheric Chemistry and Physics*, 17(21):13317–13327.

Jiang, H., Feingold, G., Jonsson, H. H., et al. (2008). Statistical comparison of properties of simulated and observed cumulus clouds in the vicinity of Houston during the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS). *J. Geophys. Res.*, 113:D13205.

Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python. http://www.scipy.org/.

Joseph, J. H. and Cahalan, R. F. (1990). Nearest neighbor spacing of fair weather cumulus clouds. *J. Appl. Meteorol.*, 29(8):793–805.

Kang, S. and Bryan, G. (2011). A Large-Eddy Simulation Study of Moist Convection Initiation over Heterogeneous Surface Fluxes. *Mon. Wea. Rev.*, 139:2901–2917.

Kirshbaum, D., Adler, B., Kalthoff, N., et al. (2018). Moist orographic convection: Physical mechanisms and links to surface-exchange processes. *Atmosphere*, 9(3):80.

Klocke, D., Brueck, M., Hohenegger, C., and Stevens, B. (2017). Rediscovery of the doldrums in storm-resolving simulations over the tropical atlantic. *Nature Geosci.*, 10(12):891–896.

Kwon, Y. C. and Hong, S.-Y. (2016). A mass-flux cumulus parameterization scheme across gray-zone resolutions. *Mon. Wea. Rev.*, 145(2):583–598.

Lamer, K. and Kollias, P. (2015). Observations of fair-weather cumuli over land: Dynamical factors controlling cloud size and cover. *J. Geophys. Res.*, 42(20):2015GL064534.

Lee, J. M., Zhang, Y., and Klein, S. A. (2019). The effect of land surface heterogeneity and background wind on shallow cumulus clouds and the transition to deeper convection. *J. Atmos. Sci.*, 76(2):401–419.

LeMone, M. A. (1973). The structure and dynamics of horizontal roll vortices in the planetary boundary layer. *J. Atmos. Sci.*, 30(6):1077–1091.

Löhnert, U., Schween, J. H., Acquistapace, C., et al. (2015). JOYCE: Jülich observatory for cloud evolution. *Bull. Amer. Meteor. Soc.*, 96(7):1157–1174.

Lohou, F. and Patton, E. (2014). Surface Energy Balance and Buoyancy Response to Shallow Cumulus Shading. *J. Atmos. Sci.*, 71:665–682.

López, R. E. (1978). Internal structure and development processes of c-scale aggregates of cumulus clouds. *Mon. Wea. Rev.*, 106(10):1488–1494.

Lotka, A. J. (1920). Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences*, 6(7):410–415.

Marke, T., Crewell, S., Schemann, V., et al. (2018). Long-term observations and high-resolution modeling of midlatitude nocturnal boundary layer processes connected to low-level jets. *J. Appl. Meteor. Climatol.*, 57(5):1155–1170.

Marke, T., Löhnert, U., Schemann, V., and Crewell, S. (2019). Detection of land surface induced atmospheric water vapor patterns. *Atmospheric Chemistry and Physics*, submitted.

Medeiros, B. and Nuijens, L. (2016). Clouds at barbados are representative of clouds across the trade wind regions in observations and climate models. *Proceedings of the National Academy of Sciences*, 113(22):E3062–E3070.

Nair, U. S., Weger, R. C., Kuo, K. S., and Welch, R. M. (1998). Clustering, randomness, and regularity in cloud fields: 5. the nature of regular cumulus cloud fields. *J. Geophys. Res.: Atmospheres*, 103(D10):11363–11380.

Neggers, R. A. J. (2015). Exploring bin-macrophysics models for moist convective transport and clouds. *J. Adv. Model Earth Syst.*, 7(4):2079–2104.

Neggers, R. A. J., Griewank, P. J., and Heus, T. (2019). Powerlaw scaling in the internal variability of cumulus cloud size distributions due to subsampling and spatial organization. *J. Atmos. Sci.*

Neggers, R. A. J., Jonker, H. J. J., and Siebesma, A. P. (2003). Size statistics of cumulus cloud populations in large-eddy simulations. *J. Atmos. Sci.*, 60(8):1060–1074.

Neggers, R. A. J. and Siebesma, A. P. (2013). Constraining a system of interacting parameterizations through multiple-parameter evaluation: Tracing a compensating error between cloud vertical structure and cloud overlap. *J. Climate*, 26(17):6698–6715.

Neggers, R. A. J., Siebesma, A. P., and Heus, T. (2012). Continuous single-column model evaluation at a permanent meteorological supersite. *Bull. Amer. Meteor. Soc.*, 93(9):1389–1400.

Nuijens, L., Medeiros, B., Sandu, I., and Ahlgrimm, M. (2015). Observed and modeled patterns of covariability between low-level cloudiness and the structure of the trade-wind layer. *J. Adv. Model Earth Syst.*, 7(4):1741–1764.

Nuijens, L., Serikov, I., Hirsch, L., Lonitz, K., and Stevens, B. (2014). The distribution and variability of low-level cloud in the north atlantic trades. *Q. J. Roy. Meteorol. Soc.*, 140(684):2364–2374.

Park, S. (2014). A unified convection scheme (UNICON). part I: Formulation. *J. Atmos. Sci.*, 71(11):3902–3930.

Patton, E., Sullivan, P., and Moeng, C. (2005). The Influence of Idealized Heterogeneity on Wet and Dry Planetary Boundary Layers Coupled to the Land Surface. *J. Atmos. Sci.*, 62:2078–2097.

Peters, O., Neelin, J. D., and Nesbitt, S. W. (2009). Mesoscale convective systems and critical clusters. *J. Atmos. Sci.*, 66(9):2913–2924.

Pielke, R. A. (2001). Influence of the spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. *Reviews of Geophysics*, 39(2):151–177.

Plank, V. G. (1969). The size distribution of cumulus clouds in representative florida populations. *J. Appl. Meteor.*, 8(1):46–67.

Plant, R. S. and Craig, G. C. (2008). A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, 65(1):87–105.

Randall, D. A. and Huffman, G. J. (1980). A stochastic model of cumulus clumping. *J. Atmos. Sci.*, 37(9):2068–2078.

Rasp, S., Selz, T., and Craig, G. C. (2018). Variability and clustering of midlatitude summertime convection: Testing the craig and cohen theory in a convection-permitting ensemble with stochastic boundary layer perturbations. *J. Atmos. Sci.*, 75(2):691–706.

Rauber, R. M., Stevens, B., Ochs, H. T., et al. (2007). Rain in shallow cumulus over the ocean: The RICO campaign. *Bulletin of the American Meteorological Society*, 88(12):1912–1928.

Reilly, S., Gesso, S. D., and Neggers, R. (2019). Configuring les in sparsely sampled areas in the subtropical atlantic. *Journal of Applied Meteorology and Climatology*, submitted.

Rieck, M., Hohenegger, C., and van Heerwaarden, C. C. (2014). The influence of land surface heterogeneities on cloud size development. *Mon. Wea. Rev.*, 142(10):3830–3846.

Robins, V. and Turner, K. (2016). Principal component analysis of persistent homology rank functions with case studies of spatial point patterns, sphere packing and colloids. *Physica D: Nonlinear Phenomena*, 334:99–117.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Sakradzija, M. and Hohenegger, C. (2017). What determines the distribution of shallow convective mass flux through a cloud base? *J. Atmos. Sci.*, 74(8):2615–2632.

Schalkwijk, J., Jonker, H. J. J., Siebesma, A. P., and Bosveld, F. C. (2015). A year-long large-eddy simulation of the weather over cabauw: An overview. *Mon. Wea. Rev.*, 143(3):828–844.

Seifert, A. and Beheng, K. D. (2004). A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description. *Meteorol. Atmos. Phys.*, 92.

Senf, F., Klocke, D., and Brueck, M. (2018). Size-resolved evaluation of simulated deep tropical convection. *Mon. Wea. Rev.*, 146(7):2161–2182.

Sengupta, S., Welch, R., Navar, M., Berendes, T., and Chen, D. (1990). Cumulus cloud field morphology and spatial patterns derived from high spatial resolution landsat imagery. *J. Appl. Meteorol.*, 29(1245-1267).

Siebesma, A. P. and Cuijpers, J. W. M. (1995). Evaluation of parametric assumptions for shallow cumulus convection. *J. Atmos. Sci.*, 52(6):650–666.

Simmons, A. (1986). Orography and the development of the ecmwf model. In *Seminar/Workshop on Observation, Theory and Modelling of Orographic effects. Seminar: 15-19 September 1986, Workshop: 19-20 September 1986*, volume 2, pages 129–164, Shinfield Park, Reading. ECMWF, ECMWF.

Simmons, A. J. and Burridge, D. M. (1981). An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. *Mon. Wea. Rev.*, 109(4):758–766.

Sommeria, G. (1976). Three-dimensional simulation of turbulent processes in an undisturbed trade wind boundary layer. *J. Atmos. Sci.*, 33.

Stein, J. (2004). Exploration of some convective regimes over the alpine orography. *Q. J. Roy. Meteorol. Soc.*, 130(597):481–502.

Stevens, B. (2005). ATMOSPHERIC MOIST CONVECTION. *Annual Review of Earth and Planetary Sciences*, 33(1):605–643.

Stevens, B. (2007). On the growth of layers of nonprecipitating cumulus convection. *J. Atmos. Sci.*, 64(8):2916–2931.

Tiedtke, M., Heckley, W. A., and Slingo, J. (1988). Tropical forecasting at ECMWF: The influence of physical parametrization on the mean structure of forecasts and analyses. *Q. J. Roy. Meteorol. Soc.*, 114(481):639–664.

Tobin, I., Bony, S., and Roca, R. (2012). Observational evidence for relationships between the degree of aggregation of deep convection, water vapor, surface fluxes, and radiation. *J. Climate*, 25(20):6885–6904.

Tompkins, A. M. and Semie, A. G. (2017). Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *J. Adv. Model Earth Syst.*, 9(2):1046–1068.

Trivej, P. and Stevens, B. (2010). The echo size distribution of precipitating shallow cumuli. *J. Atmos. Sci.*, 67(3):788–804.

van der Dussen, J. J., de Roode, S. R., Ackerman, A. S., et al. (2013). The GASS/EUCLIPSE model intercomparison of the stratocumulus transition as observed during ASTEX: LES results. *J. Adv. Model Earth Syst.*, 5.

van Heerwaarden, C. C. and de Arellano, J. V.-G. (2008). Relative Humidity as an Indicator for Cloud Formation over Heterogeneous Land Surfaces. *J. Atmos. Sci.*, 65:3263–3277.

van Laar, T. W., Schemann, V., and Neggers, R. A. J. (2019). Investigating the diurnal evolution of the cloud size distribution of continental cumulus convection using multiday LES. *J. Atmos. Sci.*, 76(3):729–747.

van Zanten, M. C., Stevens, B., Nuijens, L., et al. (2011). Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *J. Adv. Model Earth Syst.*, 3.

Vial, J., Bony, S., Dufresne, J.-L., and Roehrig, R. (2016). Coupling between lower-tropospheric convective mixing and low-level clouds: Physical mechanisms and dependence on convection scheme. *J. Adv. Model Earth Syst.*, 8(4):1892–1911.

Vogel, R., Nuijens, L., and Stevens, B. (2016). The role of precipitation and spatial organization in the response of trade-wind clouds to warming. *J. Adv. Model Earth Syst.*, 8(2):843–862.

Volterra, V. (1926). Variazioni e fluttuazioni del numero dindividui in specie animali conviventi. *Mem. Acad. Lincei Roma.*

Wagner, T. M. and Graf, H.-F. (2010). An ensemble cumulus convection parameterization with explicit cloud treatment. *J. Atmos. Sci.*, 67(12):3854–3869.

Weger, R. C., Lee, J., and Welch, R. M. (1993). Clustering, randomness, and regularity in cloud fields: 3. the nature and distribution of clusters. *J. Geophys. Res.*, 98(D10):18449–18463.

Weger, R. C., Lee, J., Zhu, T., and Welch, R. M. (1992). Clustering, randomness and regularity in cloud fields: 1. theoretical considerations. *J. Geophys. Res.*, 97(D18):20519.

White, B. A., Buchanan, A. M., Birch, C. E., Stier, P., and Pearson, K. J. (2018). Quantifying the effects of horizontal grid length and parameterized convection on the degree of convective organization using a metric of the potential for convective interaction. *J. Atmos. Sci.*, 75(2):425–450.

Wielicki, B. A. and Welch, R. M. (1986). Cumulus cloud properties derived using landsat satellite data. *J. Appl. Meteor. Climatol.*, 25(3):261–276.

Windmiller, J. (2017). *Organization of tropical convection*. PhD thesis, Ludwig–Maximilian University of Munich.

Wing, A. A. (2019). Self-aggregation of deep convection and its implications for climate. *Current Climate Change Reports*, 5(1):1–11.

Wing, A. A. and Cronin, T. W. (2015). Self-aggregation of convection in long channel geometry. *Q. J. Roy. Meteorol. Soc.*, 142(694):1–15.

Wing, A. A., Emanuel, K., Holloway, C. E., and Muller, C. (2017). Convective self-aggregation in numerical simulations: A review. *Surveys in Geophysics*, 38(6):1173–1197.

Wing, A. A. and Emanuel, K. A. (2014). Physical mechanisms controlling self-aggregation of convection in idealized numerical modeling simulations. *J. Adv. Model Earth Syst.*, 6(1):59–74.

Wing, A. A., Reed, K. A., Satoh, M., Stevens, B., Bony, S., and Ohno, T. (2018). Radiative–convective equilibrium model intercomparison project. *Geoscientific Model Development*, 11(2):793–813.

Wood, R. and Field, P. R. (2011). The distribution of cloud horizontal sizes. *J. Climate*, 24(18):4800–4816.

Wyngaard, J. C. (2004). Toward numerical modeling in the 'Terra Incognita'. *J. Atmos. Sci.*, 61(14):1816–1826.

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M. (2014). The ICON (ICOsahedral non-hydrostatic) modelling framework of DWD and MPI-m: Description of the non-hydrostatic dynamical core. *Q. J. Roy. Meteorol. Soc.*, 141(687):563–579.

Zhang, M., Bretherton, C. S., Blossey, P. N., et al. (2013). CGILS: Results from the first phase of an international project to understand the physical mechanisms of low cloud feedbacks in single column models. *J. Adv. Model Earth Syst.*, 5(4):826–842.

Zhang, Y. and Klein, S. A. (2013). Factors controlling the vertical extent of fair-weather shallow cumulus clouds over land: Investigation of diurnal-cycle observations collected at the ARM southern great plains site. *J. Atmos. Sci.*, 70(4):1297–1315.

Zhang, Y., Klein, S. A., Fan, J., et al. (2017). Large-eddy simulation of shallow cumulus over land: A composite case based on arm long-term observations at its southern great plains site. *J. Atmos. Sci.*, 74.

Zhao, G. and Di Girolamo, L. (2007). Statistics on the macrophysical properties of trade wind cumuli over the tropical western atlantic. *J. Geophys. Res.*, 112:D10204.

Zhu, T., Lee, J., Weger, R. C., and Welch, R. M. (1992). Clustering, randomness, and regularity in cloud fields: 2. cumulus cloud fields. *J. Geophys. Res.*, 97(D18):20537.

Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Roel A.J. Neggers betreut worden.

Köln, November 2019

SIGNATURE

Teilpublikationen

Thirza W. van Laar, Vera Schemann and Roel A.J. Neggers (2019), Investigating the Diurnal Evolution of the Cloud Size Distribution of Continental Cumulus Convection Using Multiday LES. *Journal of the Atmospheric Sciences*, vol. 76, no.3, doi: 10.1175/JAS-D-18-0084.1.

Thirza W. van Laar and Roel A.J. Neggers (2019), On the size dependence of cumulus cloud spacing. *Journal of Geophysical Research: Atmospheres*, submitted

Weitere Publikationen

José Luis Licón-Saláiz, Henri Riihimäki and Thirza W. van Laar (2018), Topological characterization of shallow cumulus cloud fields using persistent homology. *International Workshop on Climate Informatics*

R. Heinze et al. (2017), Large-Eddy simulations over Germany using ICON: a comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society*, vol. 143, pages 69-100, doi: 10.1002/qj.2947.

B.J.H. van Stratum et al. (2012), Study of the diurnal variability of atmospheric chemistry with respect to boundary layer dynamics during DOMINO. *Atmos. Chem. Phys. Discuss.*, **12**, 6519-6550.